DISCUSSION PAPERS
COMMUNITY AFFAIRS DEPARTMENT

# ALTERNATIVE FINANCIAL SERVICE PROVIDERS AND THE SPATIAL VOID HYPOTHESIS

Tony E. Smith
University of Pennsylvania

Marvin M. Smith
Federal Reserve Bank of Philadelphia

John Wackes
Federal Reserve Bank of Philadelphia

March 2007

FEDERAL RESERVE BANK OF PHILADELPHIA
Ten Independence Mall, Philadelphia, PA 19106-1574 • (215) 574-6458 • www.philadelphiafed.org/cca

# ALTERNATIVE FINANCIAL SERVICE PROVIDERS AND THE SPATIAL VOID HYPOTHESIS

Tony E. Smith
University of Pennsylvania

Marvin M. Smith
Federal Reserve Bank of Philadelphia

John Wackes
Federal Reserve Bank of Philadelphia

March 2007

---

# TABLE OF CONTENTS

## 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[1] 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. Research has shown that many AFSP patrons are minority and low-income consumers.[2] While this fast growing nonbank segment of the financial industry seems to provide its customers with the services they need, the price for the services is high relative to the price for 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 because mainstream financial institutions may not be present in areas where AFSPs' patrons are located. This latter hypothesis, which we call 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 on several counties in the Commonwealth of Pennsylvania. It examines whether AFSPs fill 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.

## BACKGROUND

Alternative financial service providers (AFSPs) have been around in various forms for some time. According to one account, in 1986 there were about 2,200 alternative providers nationwide, but that number had grown to more than 10,000 by 1994.[4] The total number of AFSPs has greatly increased since then: The number of pawnshops alone has been estimated to be between 12,000 and 14,000, while in 2004 the number of payday lending outlets was estimated to be 22,000.[5] Just in Pennsylvania, the number of licensed check cashers and pawnbrokers together grew 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

---

[1] General Accounting Office, "Electronic Transfers," Report to the Subcommittee on Oversight and Investigations, House of Representatives, GAO-02-913, 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* (2001).

[3] See Noah Sawyer and Kenneth Temkin, *Analysis of Alternative Financial Service Providers* (The Fannie Mae Foundation, 2004).

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

[5] See James H. Carr and Jenny Schuetz, *Financial Services in Distressed Communities: Framing the Issue, Finding Solutions* (The Fannie Mae Foundation, 2001); and Mark Flannery and Katherine Samolyk, "Payday Lending: Do the Costs Justify the Price?" Working Paper 2005-09, FDIC Center for Financial Research (2005).

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 those for 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 because of the absence of mainstream financial institutions, or (iii) whether they offer more convenient hours and a host of nonfinancial services in addition to basic financial needs. The FMF study dealt with the first two of these areas of interest.[8] While this study will formally address the first two, it will also shed some light on the third.

## 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 profitably 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), they might also be thought to be acting pragmatically.

It is understandable why such consumers might use some AFSPs for their credit needs, but some people might question why consumers who have accounts at traditional financial institutions would use the services of AFSPs.[9] 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 nonfinancial services and products such as postage stamps,

[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 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 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.

[9] However, the use of AFSPs by those with bank accounts should not be too surprising. In fact, a study by Seidman, Hababou, and Kramer found that 26.1 percent of the low- and moderate-income consumers they surveyed in Los Angeles, Chicago, and Washington, D.C. who had a bank account used a nonbank to cash their checks primarily because of convenience.

train and bus passes, notary services, lottery tickets, payment of utility bills, and prepaid telephone cards.[10]

Finding a comprehensive rationale as to why AFSPs might be complements to traditional financial institutions is beyond the scope of this study. Instead, we will investigate the spatial void hypothesis as it relates to selected counties in the Commonwealth of Pennsylvania.[11]

## METHODOLOGY

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.[12,13] In this study, "banks" will refer to all full-service offices of a banking organization, including banks' head offices or their branches, if any. The analysis proceeds in several stages. First, following the FMF study, the spatial clustering of AFSPs and banks is 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."[14] 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 AFSPs are serving markets with significantly lower income levels than those served by 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, by asking whether the residents of each given neighborhood (block-group) have significantly greater access to banks than would be expected if AFSPs and banks were indistinguishable.

An important feature of this 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

---

[10] See the paper by John Caskey, p. 3.

[11] Although this study concentrates on the location of AFSPs vis-à-vis banks, 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) (2004), pp. 191-205; and Steven M. Graves, "Landscapes of Prediction, Landscapes of Neglect: A Location Analysis of Payday Lenders and Banks," *The Professional Geographer,* 55(3) (2003), pp. 303-17.

[12] While all of the counties in Pennsylvania had banks, 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 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.

[13] In this analysis, AFSPs are composed of only check cashers and pawnbrokers.

[14] FMF study, p.7.

state licensing data as of May 2005 from Pennsylvania's Department of Banking. The Department of Banking's address location file provided information for the two types of AFSPs of interest in this analysis — check cashers and pawnbrokers. The data reflect 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."[15] Similarly, the Pawnbrokers License 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."[16]

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

We perform the analysis of the demographic make-up of block-groups where AFSPs and banks are located 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.

Finally, the addresses of the banks/bank branches were obtained from the Federal Deposit Insurance Corporation's (FDIC) database, which contains the locations of all FDIC-insured full-service bank branches in the Commonwealth of Pennsylvania.[17]

## RESULTS

The four Pennsylvania counties in this study had a total of 333 AFSPs, with nearly 70 percent of them in Philadelphia County (Table 1). Alternatively, all four counties had 1,339 banks, with all but one county having 320 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 |

---

[15] *Pennsylvania Bulletin*, 28(18) ( May 2, 1998), Harrisburg, PA.

[16] Pawnbrokers License Act of 1937, P.L. 200, No. 51.

[17] 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.

## Individual Cluster Analyses of AFSPs and Banks

In performing the separate cluster analyses of AFSPs and banks, our 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 Ripley's[18] *K-function* approach, which allows a systematic examination of clustering at alternative spatial scales. A second and even more important departure is the introduction of population density as a 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, we mapped the locations of AFSPs and banks using the geocoding procedure in the ArcMap software.[19] The point locations of AFSPs and banks are shown for Philadelphia County in Figures 1a and 1b. 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 of Crime Stat 2.0 software, as detailed in the manual for this software.[20] For our purposes, it

**FIGURE 1A**
**PHILADELPHIA AFSPs**

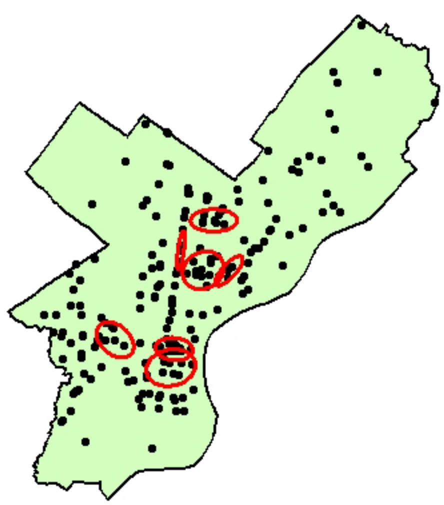
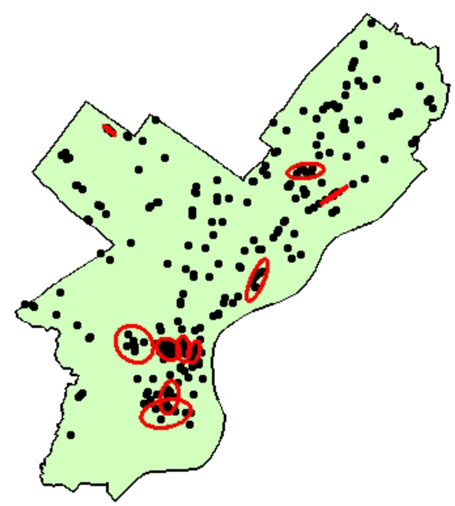


**FIGURE 1B**
**PHILADELPHIA BANKS**



[18] See B.D Ripley, "The Second-order Analysis of Stationary Point Patterns," *Journal of Applied Probability*, 13 (1976), pp. 255-66.

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

[20] This manual is obtainable online at http://www.mappingcrime.org/crimestat.htm. A full description of the nearest neighbor hierarchical clustering procedure used here starts on page 216 of Chapter 3.

is enough to say that the underlying idea of this procedure is essentially 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 "admissible" cluster. As in the FMF study, these values were 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.[21]

While these regions do appear to correspond to the regions of highest concentration in each pattern,[22] 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). 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 Figures 3a and 4a, which are enlargements of Figures 1a and 1b, respectively. Here the green background represents population densities in each of the block-groups shown on the map.[23] 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, we develop a second procedure that 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 the area of each block-group, in our approach we postulate that these likelihoods are proportional to population.[24] 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 corresponds to one of the AFSPs near the western edge of Philadelphia County.

---

[21] 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 (or vertical) 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, we set this parameter to its minimum value of one standard deviation.

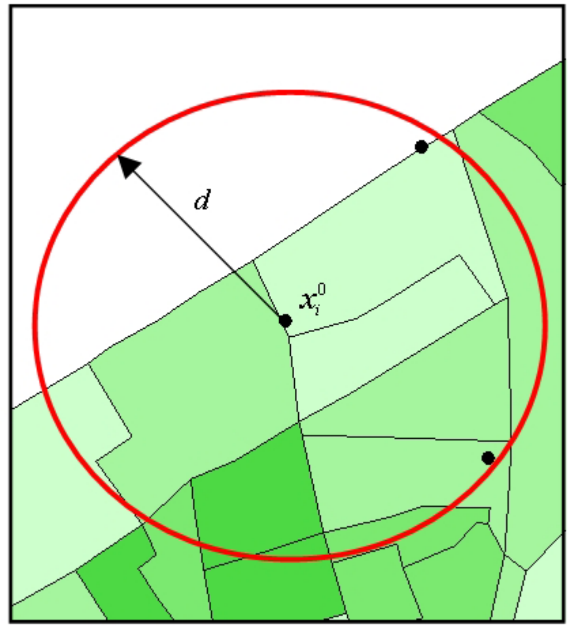
[22] Note that many establishments are so close together that they do not register as separate dots in these figures; so simple visual inspections of the patterns are necessarily limited by this fact.

[23] The block-groups left blank were excluded from the analysis for reasons including zero population levels. In many instances they are areas with parks or other uses.

[24] Note that a similar "risk-adjusted" version of the NNHC procedure is available in Crime Stat 2.0 (starting on p. 235 of Chapter 3). However, this procedure is considerably more complex than our 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.

Here $d$ equals one-half mile, and the figure shows that $K_i^0(d) = 2$.[25] 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), we can employ Monte Carlo methods to estimate the sampling distributions of such point counts. To do so, we need only simulate a large number of replicate point patterns, $X^{(s)} = \{x_i^{(s)} : i = 1,..,n-1\}, s = 1,..,N$, of size $n-1$, from a probability distribution proportional to population.[26] 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 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, 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

**FIGURE 2**
**POINT COUNTS FOR PATTERNS**




$$(1) \quad P_i(d) = \frac{M_i(d)}{N+1}$$

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.[27] So, for example, if $M_i(d) = 10$, $P_i(d) = 10/1001 < .01$ would imply that the estimated chance of observing a value as large as $K_i^0(d)$ is less than one in 100. In this case $K_i^0(d)$ might indeed be regarded as "unusually large" after accounting for population.[28]

These p-values can be mapped, and they provide a clear visual picture of where counts are unusually high. To

---

[25] 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 (1995), pp. 93-116.

[26] Technically this amounts to a two-stage sampling procedure. If $p_i$ denotes the fraction of population in block-group i, 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). 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.

[27] One of the authors programmed this simulation procedure and all others discussed in this paper in MATLAB.

[28] We should note here that our approach overcomes one additional limitation of both the NNHC procedure and its risk-adjusted version described in footnote 20. This limitation is well illustrated by the example in Figure 2, 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 our Monte Carlo approach are subject to the same edge effects, the p-values obtained are not biased by this effect.

compare our results with those of the NNHC procedure described above, we have again selected a single p-value threshold of $p = .01$. 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. 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.[29] With this definition, the points inside

**FIGURE 3A**
**PHILADELPHIA AFSPs (NNHC)**

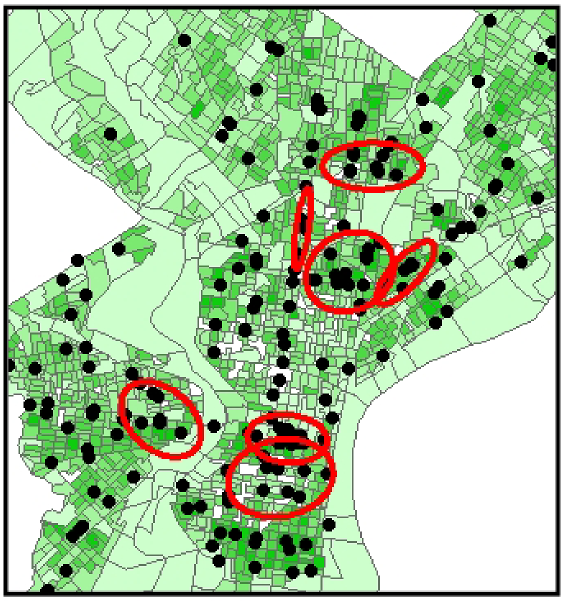


**FIGURE 3B**
**PHILADELPHIA AFSPs (K-FUNCTION)**

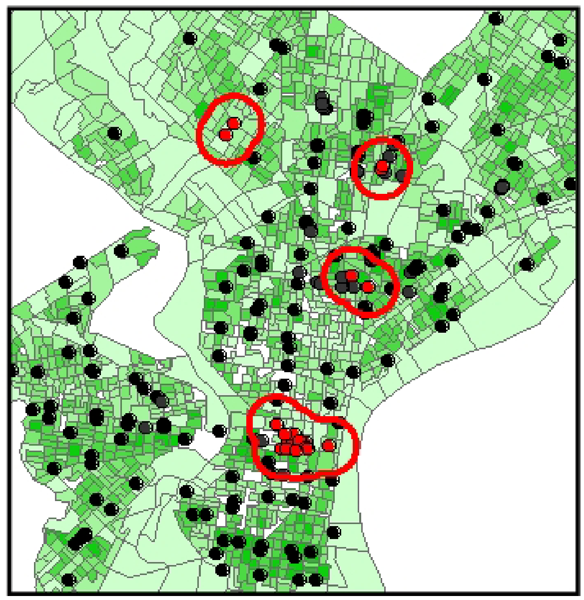


**FIGURE 4A**
**PHILADELPHIA BANKS (NNHC)**

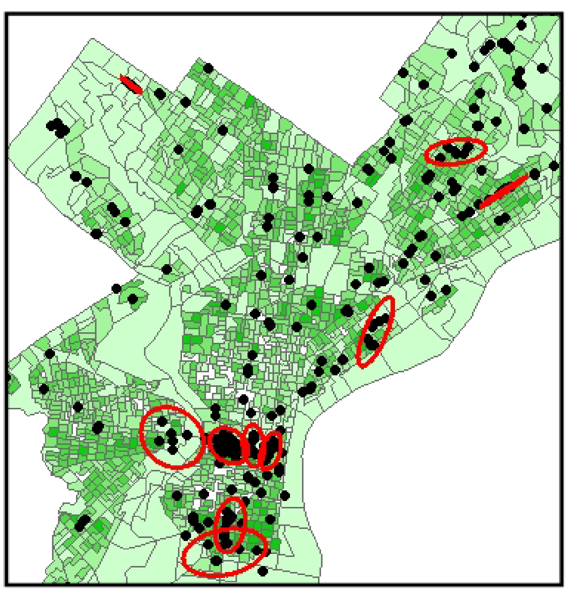
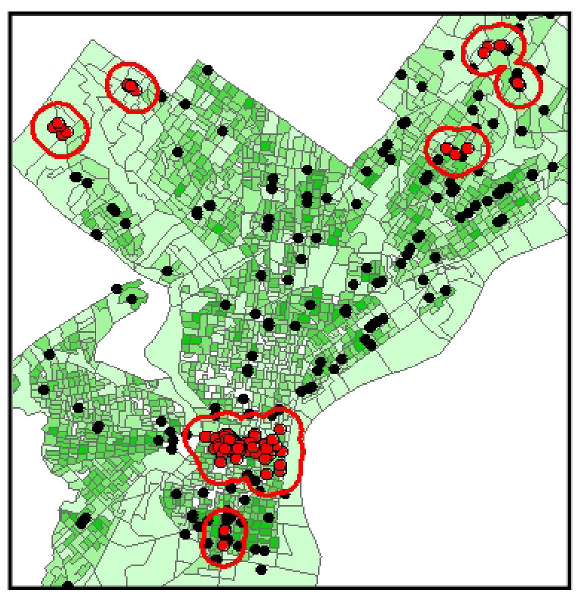


**FIGURE 4B**
**PHILADELPHIA BANKS (K-FUNCTION)**



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

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.

Before comparing these two procedures, we should 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. It should be clear that a range of other values can be considered. In fact, this has been done, and we can report 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 our study as well. For example, choices of cluster sizes smaller than five tend to yield many small, isolated clusters. Moreover, values larger than five 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 five 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, $p = .01$, we can make observations 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), [30] who were among the first to use this method, each such p-value is valid only 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, these correlated groupings serve mainly to identify areas of potentially significant clustering.[31]

Finally, turning to the radial-distance threshold, *d*, it should be clear from the 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.[32] 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 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

---

[30] See G. Rushton and P. Lolonis, "Exploratory Spatial Analysis of Birth Defect Rates in an Urban Population," *Statistics in Medicine*, 15 (1996), pp. 717-26.

[31] We should also mention 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*, 154 (1991), pp. 327-33; and the SaTScan procedure of M. Kulldorff, "A Spatial Scan Statistic," *Communications in Statistics: Theory and Methods*, 26 (1997), pp. 1487-96. 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 (2006), pp. 180-208.

[32] 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.

methods. Overall, the K-function approach generally yields fewer bank clusters than the NNHC method as well as fewer AFSP clusters in Philadelphia County (Table 2). But this result masks other worthwhile findings. Turning first to a comparison of AFSP clusters in Figures 3a and 3b, 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.[33]

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

| | NUMBER OF AFSP 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 |

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, we 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 – 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). 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 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 daytime population levels are more accurately reflected by employment data than residential data. 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 K-function approach. What is needed is a more accurate reference measure

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

of relevant *consumer* population densities. So 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. Again, there are many approaches we could take. 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 bank is analyzed.[34] Using this approach, we 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.[35] 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 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 $j$.[36] This *individual accessibility function*, $a(d_{ij})$, is assumed here to have the following quadratic (kernel) form,

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

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 distance at which institution $i$ can expect to draw customers, as illustrated for $b = 0.5$ (one-half mile) in Figure 5.[37] The role of this parameter will be discussed further below.

---

[34] 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 (2004), pp. 315-31. See also the many references cited in that paper.

[35] For further discussion of this "indistinguishability" concept, see footnote 39.

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

[37] This, of course, serves only as a rough approximation to actual customer potential and (as in footnote 32) should ideally be estimated on the basis of actual market data.

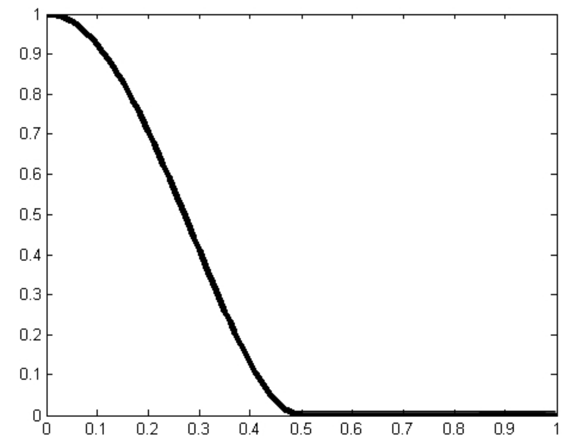
Given this individual accessibility function, the expected number of potential $i$-customers from $j$ is the product of $a(d_{ij})$ and the population, $n_j$, of block-group $j$. The probability that any potential $i$-customer is from $j$ is given by

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

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*[38] by

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

FIGURE 5
INDIVIDUAL ACCESSIBILITY FUNCTION




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. Our 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),[39] the key question of interest here is whether these incomes are higher for potential customers of banks than of AFSPs. If so, this finding would lend support to the spatial void hypothesis. To answer this question, we proceed by postulating the null hypothesis that typical income levels of potential customers are the same for banks and AFSPs, and then we test this hypothesis using Monte Carlo methods.

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 $L_{AFSP} = \{i : i = 1,..,n_{AFSP}\}$,[40] (maximum likelihood) estimates of these expected values are given respectively by

$$(5)\quad Y_B^0 = \frac{1}{n_B}\sum_{i\in L_B} Y_i^0 \qquad (6)\quad Y_{AFSP}^0 = \frac{1}{n_{AFSP}}\sum_{i\in L_{AFSP}} Y_i^0$$

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, we could

---

[38] 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. 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*.

[39] We should also note 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 (1992), pp. 189-207.

[40] 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.

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 as the given street network and local zoning restrictions). 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,

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

as given, and to ask what expected income differences would look like if the location behavior of banks and AFSPs were completely indistinguishable. "Indistinguishable" means that the specific locations called banks are simply one of many equally likely choices of $n_B$ sites from $L$.[41] Since each choice of these sites amounts to a random relabeling 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 $N$ and relabeling and recomputing 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, $\pi$, the estimates corresponding to (5) and (6) above are now given respectively by:

$$(8)\quad Y_B^\pi = \frac{1}{n_B}\sum_{i\in L_B} Y_i^\pi \qquad (9)\quad Y_{AFSP}^\pi = \frac{1}{n_{AFSP}}\sum_{i\in L_{AFSP}} Y_i^\pi$$

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, $\pi_k$, $k = 1,..,N$,[42] 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 $M_\Delta$, as a parallel to (1) above, the chance of observing a difference as large as $\Delta(0)$ under the null hypothesis of indistinguishable institutions is estimated to be

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

Hence $P_\Delta$ 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, $P_\Delta = 10/1001 < .01$ 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 100 under this null hypothesis.

Before discussing the results of these tests, we should emphasize again that such tests implicitly rely on the

---

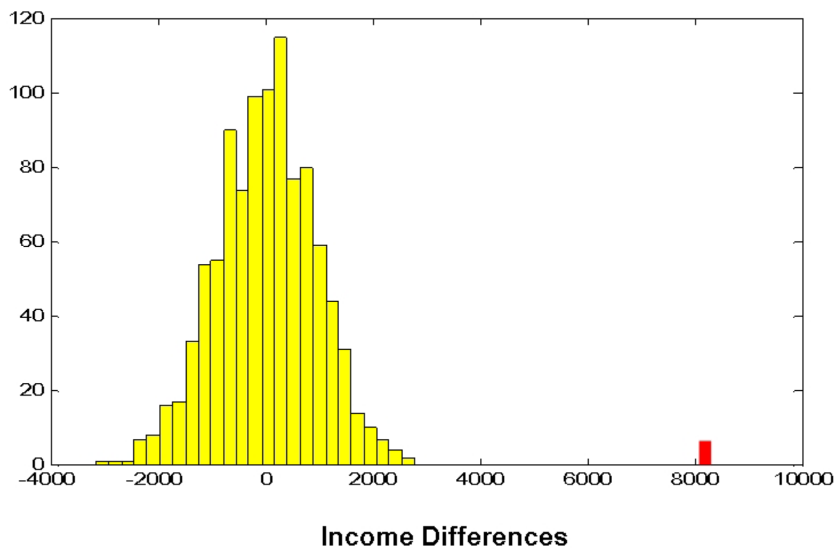
[41] Statistically, the labels of locations are then referred to as exchangeable events. From a behavioral viewpoint, this is similar to Debreu's well-known example of choosing "red buses" versus "blue buses." See G. Debreu, "Review of R.D. Luce, Individual Choice Behavior: A Theoretical Analysis," *American Economic Review*, 50 (1960), pp. 186-88.

[42] W should note here 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 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, 1995).

specific parameterization of the individual accessibility function in (2) above. Moreover (as mentioned in footnote 32), this behavioral function is generally not directly observable. However, many researchers have observed that the key parameter determining the *statistical* behavior of such kernel functions is their extent (also called their bandwidth).[43] 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, we can argue that this robustness lends considerable support to the test conclusions. To determine robustness in the present case, we used a range of extent values from one-quarter mile up to one-and-a-half miles.[44]

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 6 shows the histogram of values obtained for the observed difference, $\Delta(0)$, together with $N = 1000$ simulated differences, $[\Delta(\pi_1),..,\Delta(\pi_N)]$, 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 $8000, as shown by the red bar in the figure. The results of 1000 simulations (random relabelings) 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. On the basis of these tests we can conclude that typical customer incomes (as we have defined them) are significantly higher for banks than for AFSPs.

**FIGURE 6**
**EXPECTED INCOME DIFFERENCES**
**FOR PHILADELPHIA**




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

While the testing procedure above 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. The objective of our approach is to add a *spatial dimension* to the analysis above. To do so,

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

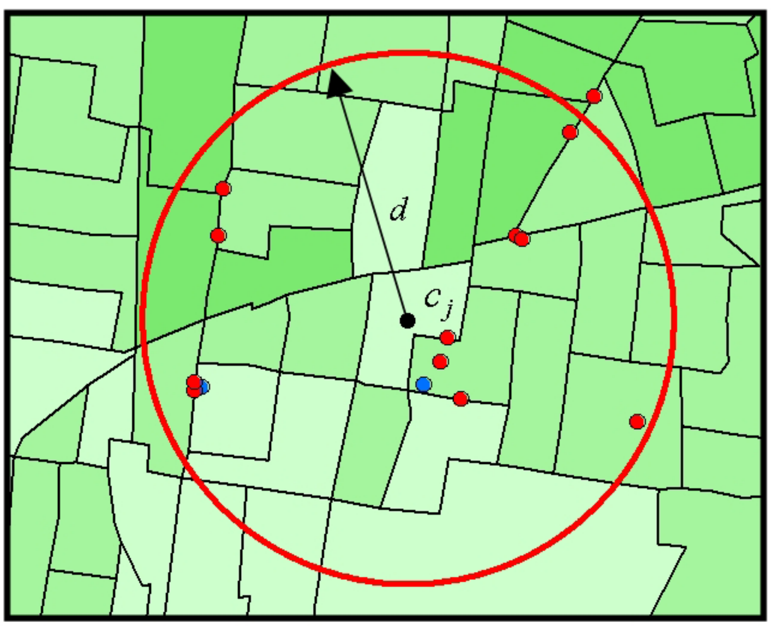
[44] In units of miles, the actual extent values used were $b = [.25,.50,.75,1.00,1.50]$.

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 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,[45] we choose to develop a simpler approach that yields exact probability calculations and avoids the need for Monte Carlo methods.

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 example is shown in Figure 7, where 13 financial institutions are seen to be within one-half mile ($d = 0.5$) of $c_j$.[46] In this case there are 11 AFSPs (red) and only two banks (blue). 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.

**FIGURE 7**
**INSTITUTION COUNTS FOR BLOCK-GROUPS**




To test this conjecture, we again appeal to the indistinguishability hypothesis and suppose that this assignment of banks and AFSPs to locations in $L$ is only one among many possible equally likely labelings 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), it is well known that the chance that $k$ of these $m$ institutions will be AFSPs is given by the hypergeometric probability:[47]

---

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

[46] The actual location of this block-group is shown in Figure 8.

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

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

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

In our case, this cumulative probability, $P_j(d)$, 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 7 (where $d = 0.5$),

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

If banks and AFSPs were, in fact, indistinguishable, the chance of such an extreme concentration of AFSPs would be extremely small. On this basis, we may conclude that this block-group has an unusually large supply of AFSPs relative to banks. 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 we must consider a range of extents ($d$ values) in order to gauge 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 that in Figures 3b and 4b, to yield a picture of "relative concentration of hot spots." We choose to show a broader range of p-values than the simple ".01-threshold" value in Figures 3b and 4b. In Figure 8, all p-values $P_j(d) \leq .10$ are shown; the darkest p-values are those where the most significant relative concentration of AFSPs occurs.[48] 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).[49] For example, the case shown in Figure 8 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: A significant clustering of AFSPs is seen in Figure 3b but no significant clustering of banks is seen in Figure 4b. At the

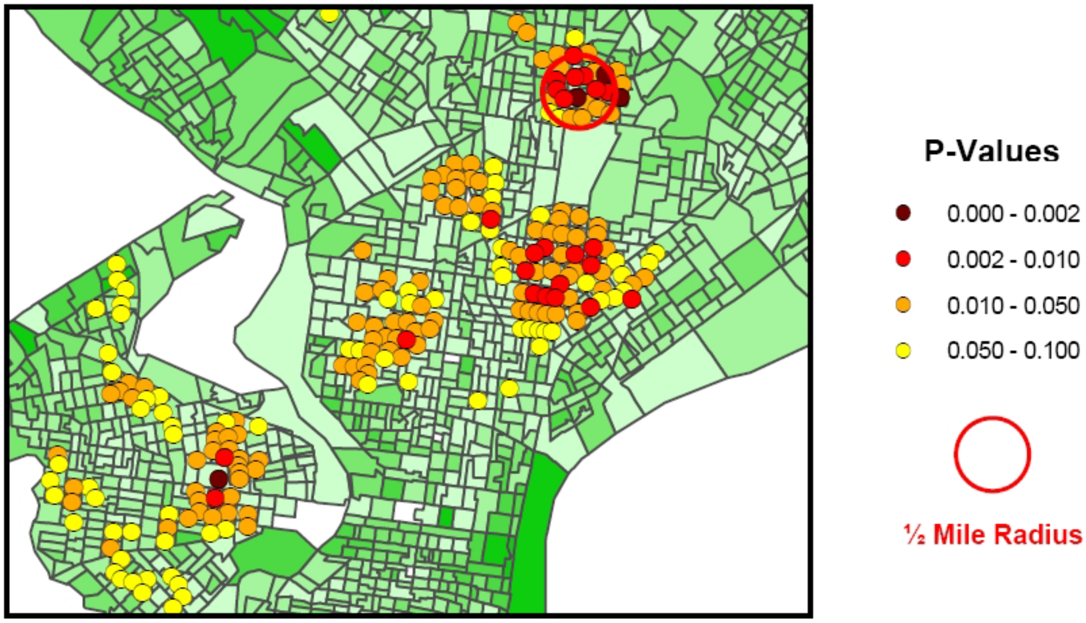
---

[47] 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 with banks. A recent discussion of this approximation issue can be found in F. Lopez-Blazquez and B. Salamanca Mino, "Binomial Approximation to Hypergeometric Probabilities," *Journal of Statistical Planning and Inference*, 87 (2000), pp. 21-29.

[48] 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."

[49] An illustration of two alternative scales is shown for banks in Figure 9.

FIGURE 8
RELATIVE CONCENTRATIONS OF AFSPs VERSUS BANKS



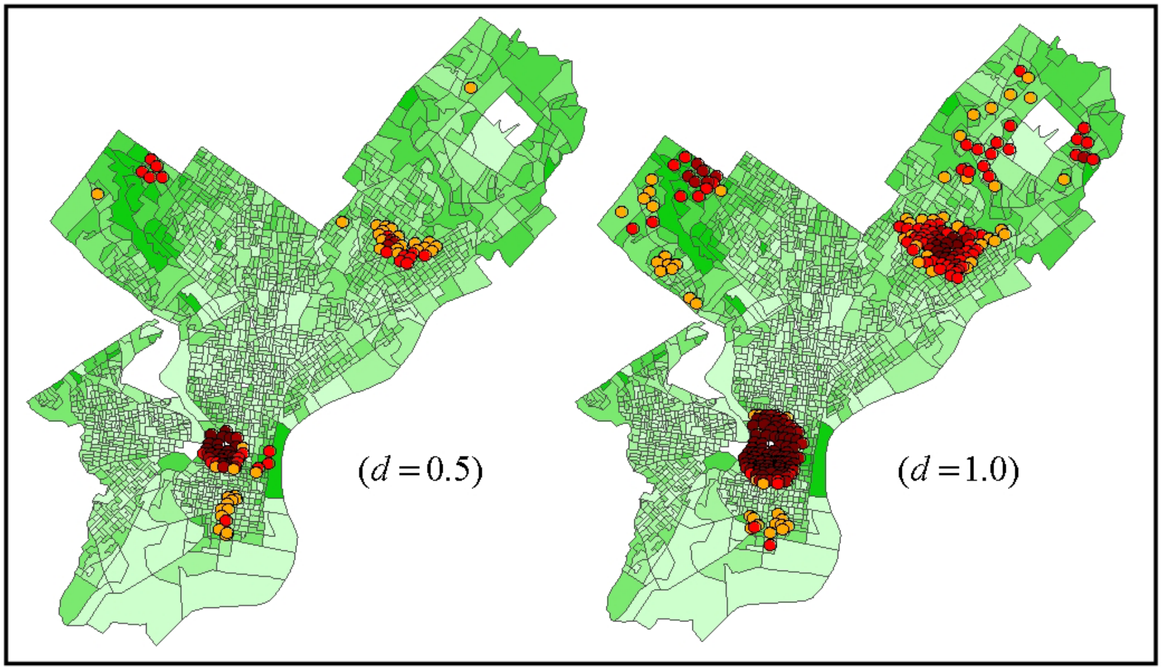

other extreme, it is interesting to note that while there was no significant clustering of either AFSPs or banks in West Philadelphia, there is a significant relative concentration of AFSPs precisely in the area where significant clustering of AFSPs was indicated by the NNHC approach in Figure 3a.

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 with banks is quite significant.[50] Finally, note that while Figures 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 with 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 9 (using the same p-value scale as in Figure 8). Results for a half-mile radius ($d = 0.5$) are shown on the left; 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 that in Figure 8, the most striking feature is the significantly high relative concentration of banks versus AFSPs in Center City Philadelphia. Of course, this is not surprising, given the nature of central business districts. But notice that this result does add information to Figure 8 – since, in principle, it is possible that neither banks nor AFSPs would exhibit significant relative concentration when one is compared with the other. Observe next that 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

[50] In Figure 4a 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 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.

FIGURE 9
RELATIVE CONCENTRATIONS OF BANKS




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 slightly more than one in five.[51] Note that while the results for a one-mile radius are 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 one-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. We 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 6, we should consider a Monte Carlo test based on an appropriate modification of the *indistinguishability hypothesis*. Again, we 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 $BG = \{j : j = 1,..,n\}$, in terms of (12) with $d = 0.5$, we first identify the subset

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

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) $$\overline{Y}_0 = \frac{1}{n_0}\sum\nolimits_{j \in BG_0} y_j \ ,$$

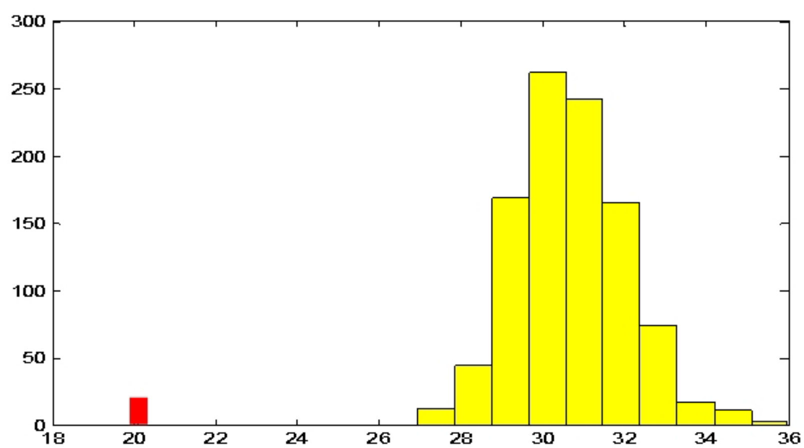
[51] In terms of the notation in (12), the exact probability is $\text{Prob}(C \geq 5 \mid 6, 321, 551) = .206$.

the appropriate null hypothesis for our 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 this case, if we draw N independent random samples of size $n_0$ from BG and denote their realized average median incomes by $\overline{Y}_k, k = 1,..,N$, 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

$$P_0 = \frac{M_0}{N+1} \tag{16}$$

As in the case of Figure 6, 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 10, where the red bar denotes the observed value, $\overline{Y}_0 = \$20,107$ and the yellow bars represent the 1000 simulated $\overline{Y}$ 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. This result provides further strong support for the spatial void hypothesis in Philadelphia County.

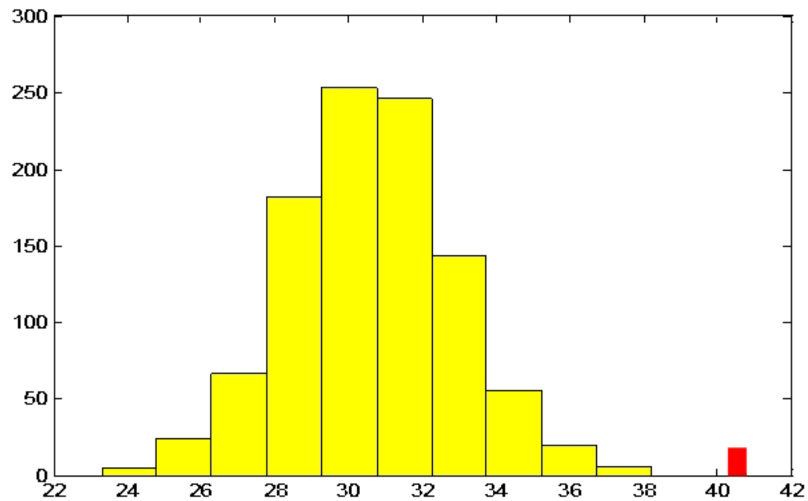
**FIGURE 10**
**INCOME TEST RESULTS FOR AFSPs**




We can perform exactly the same analysis 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. 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 $\overline{Y}_0 = \$40,796$. Again, the results for this case are extreme and are best shown by a histogram of values, as in Figure 11. Thus, the average medium income is 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 6 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

## FIGURE 11
## INCOME TEST RESULTS FOR BANKS




would, of course, yield dramatic results, as we have seen. But the present approach provides useful additional information: namely, that $\overline{Y}_0$ for AFSPs is, in fact, significantly *lower* than typical median incomes in Philadelphia and that $\overline{Y}_0$ for banks is significantly *higher*.

*Location of AFSPs and Median Income P-Values*

Table 3 shows a comparison of median incomes of block-groups with significant relative access to AFSP 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 with 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. The last column reveals that compared with 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 shows that the number of AFSP clusters of block-groups among the four counties range from 14 in Allegheny County 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 1,107 block-groups with significantly more access to AFSPs than to banks.

Note that we have not included a parallel set of results for banks (as suggested by Figure 11). The reason is that there are so few AFSPs in counties other than Philadelphia that *no* block-groups exhibit significantly high relative access to banks (at any scale of analysis) in the other three counties. 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 AFSPs

## TABLE 3
## MEDIAN INCOME OF BLOCK-GROUPS

| | TOTAL NUMBER OF BLOCK-GROUPS | AFSP CLUSTERS | AFSP MEDIAN INCOME | COUNTY MEDIAN INCOME | P-VALUE |
|---|---|---|---|---|---|
| Philadelphia County | 1,816 | 28 | $19,900 | $30,686 | <.0001 |
| Delaware County | 462 | 44 | $35,426 | $63,347 | <.0001 |
| Allegheny County | 1,107 | 14 | $34,592 | $49,988 | 0.0008 |
| Montgomery County | 543 | 35 | $40,078 | $75,830 | <.0001 |

in the county. 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, another question often raised is whether they are located predominantly in minority areas. We examine this question by considering the location patterns of AFSPs in our four study counties.

*Location of AFSPs and Minorities*

Figure 12 shows pie charts comparing 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, we can determine whether minorities are disproportionately represented in the designated block-groups. For example, minorities comprise 60.3 percent of the population in Philadelphia County, but they make up 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 overrepresented in block-groups with significant relative access to AFSP clusters in Montgomery, Delaware, and Allegheny counties but underrepresented in Philadelphia County (where they comprise 46.1 percent in the county and 34.2 percent in the designated block-groups).[52] However, there is uniformity in the overrepresentation of Hispanics in the pertinent block-groups in all of the counties.

The representation of Asians 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 overrepresentation of Hispanics more than offsets the underrepresentation of African Americans, thus resulting in an overall overrepresentation of minorities as mentioned earlier. [53]

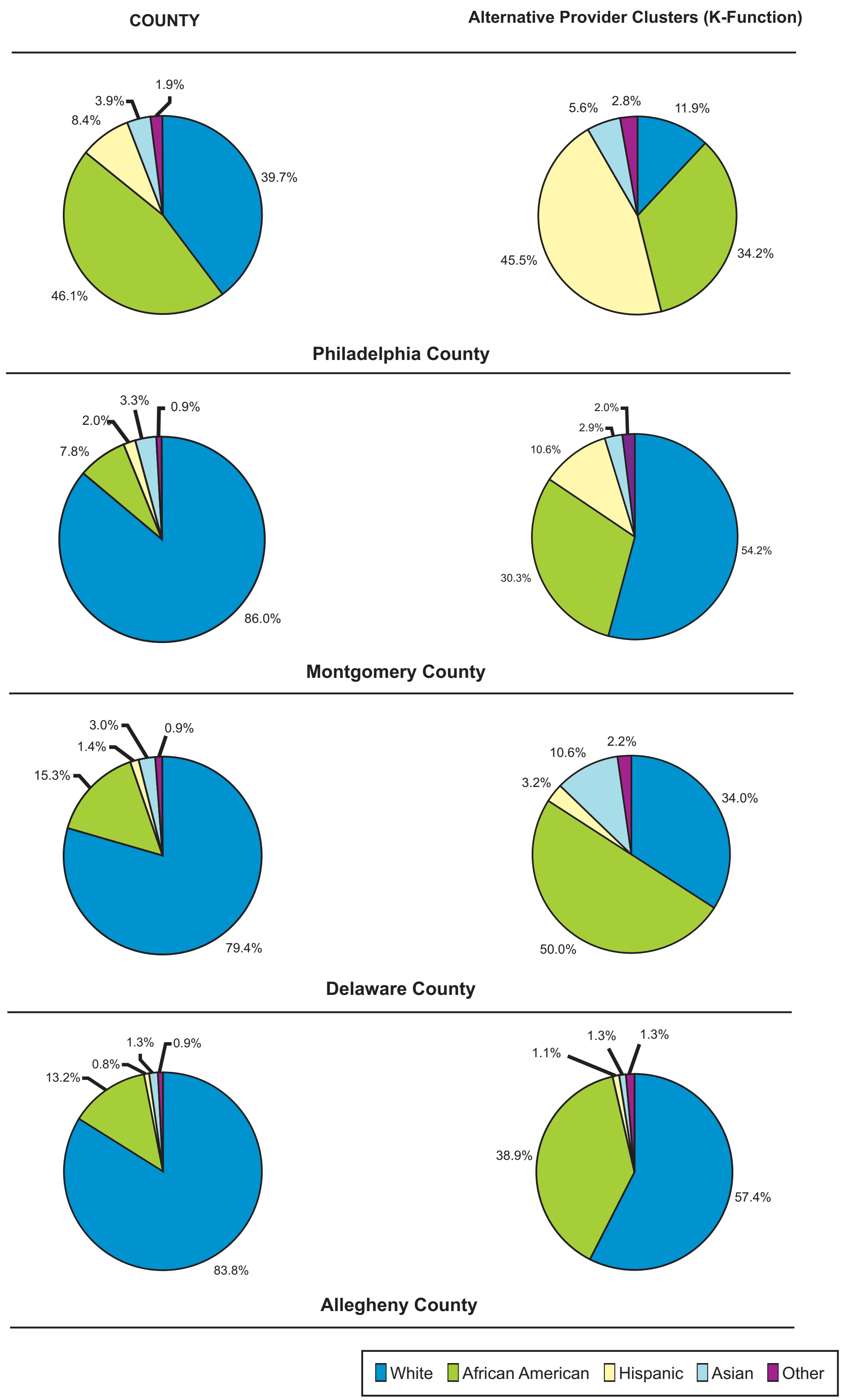
In the more suburban counties, there is a lack of overlap of clusters of banks and alternative financial service providers. As shown in the maps in the Appendix, with the exception of Center City Philadelphia and the Pittsburgh business district, there is a lack of overlap of the two types of clusters in the counties studied. Both Center City Philadelphia and downtown Pittsburgh exhibit clusters of bank branches and alternative financial service providers. The rest of the county areas show a lack of overlap.

Montgomery County contains bank branch clusters in its western portion with a cluster of alternative financial service providers in Norristown. Likewise, Delaware County exhibits a lack of overlap with clusters of alternative financial service providers in the Upper Darby area close to the county border with Philadelphia. In Delaware County, bank branch clusters are to be found in the center and in the western portion of the county. Allegheny

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

[53] These results were again confirmed using the method in footnote 52.

## FIGURE 12
## RACIAL OR ETHNIC COMPOSITION OF BLOCK-GROUPS




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

County contains bank branch clusters outside of and surrounding the city of Pittsburgh with an overlap of clusters of alternative financial service providers only in downtown Pittsburgh.

## SUMMARY AND CONCLUSIONS

The increased prominence of alternative financial service providers (AFSPs) in 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 studies that have documented that many minority and low-income households rely heavily on the services of AFSPs, these consumers might forgo valuable wealth-building opportunities. Our study has focused on (i) whether the high patronage of this nonbank 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. We carried out this investigation by focusing on four counties in the Commonwealth of Pennsylvania: Philadelphia, Delaware, Allegheny, and Montgomery.[54]

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 our study we showed that because the NNHC approach fails to account for population concentrations (and also ignores "edge effects"), it can sometimes produce misleading results. Hence, an alternative approach (based on Ripley's *K-function*) was proposed that 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.

To test the spatial void hypothesis with respect to these four study counties, we 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 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 overrepresentation 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.

---

[54] Although we focused only on four out of the 67 counties in Pennsylvania, this does not mean that AFSPs vs. banks may not be an issue in the remaining counties. The techniques discussed here can best address the spatial void hypothesis in geographical areas (in this case counties) that have sufficient clusters of AFSPs and banks.

## REFERENCES


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

Besag, J., and J. Newell. "The Detection of Clusters in Rare Diseases," *Journal of the Royal Statistical Society*, 154 (1991), pp. 327-33.

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

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

Caskey, John. "Bringing Unbanked Households into the Banking System," *Capital Xchange*, The Brookings Institution (January 2002).

Caskey, John. *Fringe Banking: Check Cashing Outlets, Pawnshops, and the Poor*. 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 (2006), pp. 180-208.

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

Debreu, G. "Review of R. D. Luce, Individual Choice Behavior: A Theoretical Analysis," *American Economic Review,* 50 (1960), pp. 186-88.

Edgington, E.S. *Radomization Tests*. Marcel Dekker, 1995.

Flannery, Mark, and Katherine Samolyk. "Payday Lending: Do the Costs Justify the Price?" Working Paper, 2005-09, FDIC Center for Financial Research (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 (1992), pp. 189-207.

Goldman, Gerald, and James R. Wells, Jr., *Check Cashers Are Good Bank Customers*. 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 (2003), pp. 303-17. http://www.mappingcrime.org/crimestat.htm.

Kulldorff, M. "A Spatial Scan Statistic," *Communications in Statistics: Theory and Methods*, 26 (1997), pp. 1487-96.

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

Pawnbrokers License Act of 1937, P.L. 200, No. 51.

*Pennsylvania Bulletin*, 28:18 (May 2, 1998), Harrisburg, PA.

Ripley, B. D. "The Second-order Analysis of Stationary Point Patterns," *Journal of Applied Probability*, 13 (1976), pp. 255-66.

Rushton, G., and P. Lolonis. "Exploratory Spatial Analysis of Birth Defect Rates in an Urban Population," *Statistics in Medicine*, 15 (1996), pp. 717-26.

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

Seidman, Ellen, Moez Hababou, and Jennifer Kramer. "A Financial Services Survey of Low- and Moderate-Income Households," presented at the Federal Reserve System Community Affairs Research Conference, Washington, D.C., April 7, 2005. http://www.chicagofed.org/cedric/promises_pitfalls_2005_conference_session2.cfm

Silverman, B.W. "Choosing a Window Width When Estimating a Density," *Biometrika*, 65 (1978), pp. 1-11.

Smith, T.E. "A Scale-sensitive Test of Attraction and Repulsion between Spatial Point Patterns," *Geographical Analysis*, 36 (2004), pp. 315-31.

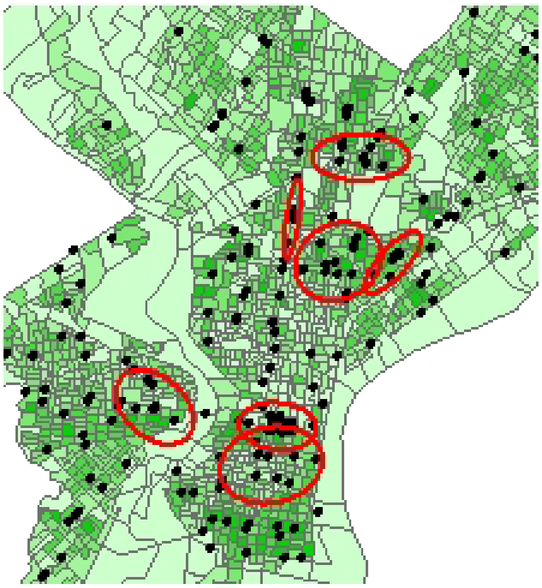
Stegman, Michael. "Banking the Unbanked: Untapped Market Opportunities for North Carolina's Financial Institutions," *Journal of the University of North Carolina School of Law* (2001).

# APPENDIX

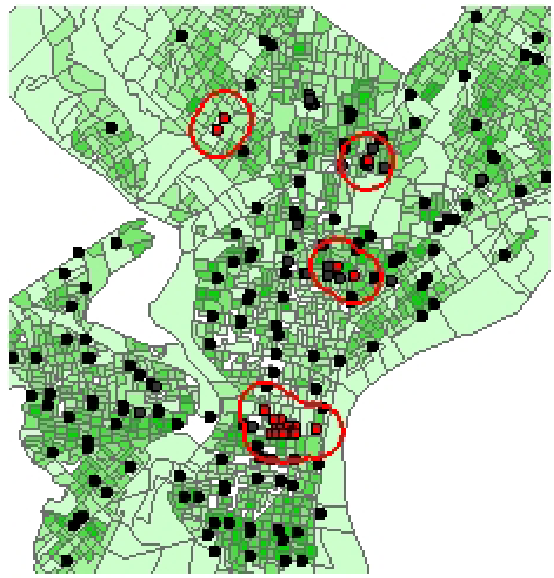
This appendix contains side-by-side figures comparing the estimation of clusters for alternative financial service providers (AFSPs) and banks in each of the study counties using nearest neighbor hierarchical clustering (NNHC) and K-function procedures.

● P-Value <= .01

**Philadelphia AFSP Clusters**

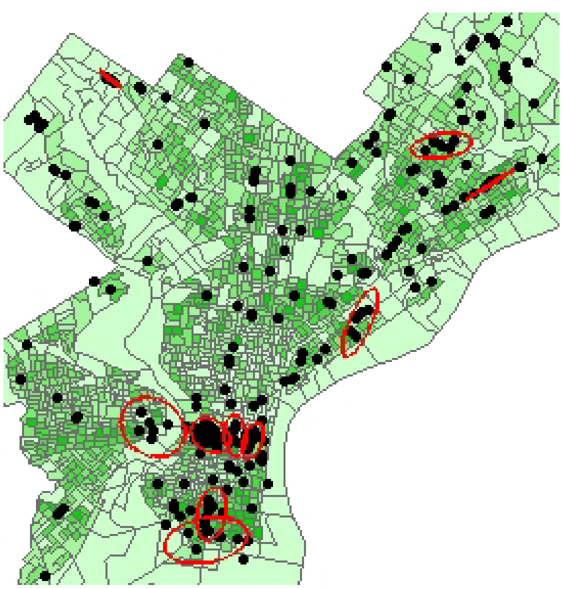


NNHC
(P-value = .01, Count >=5)

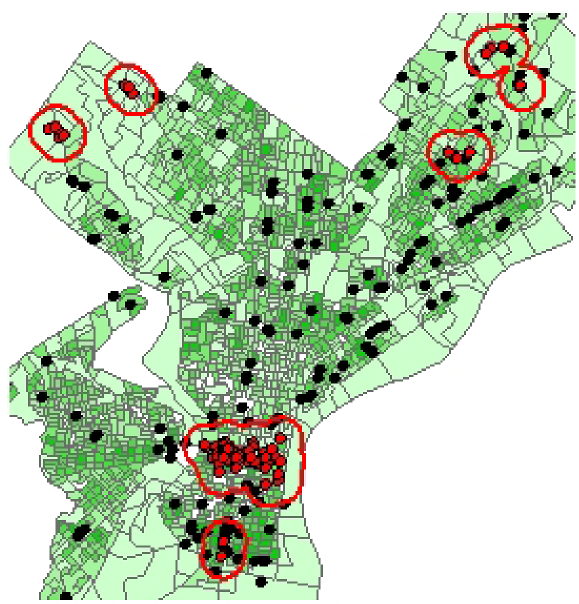


K-Function
(One-Half Mile radius, Count >=5)

**Philadelphia Bank Clusters**



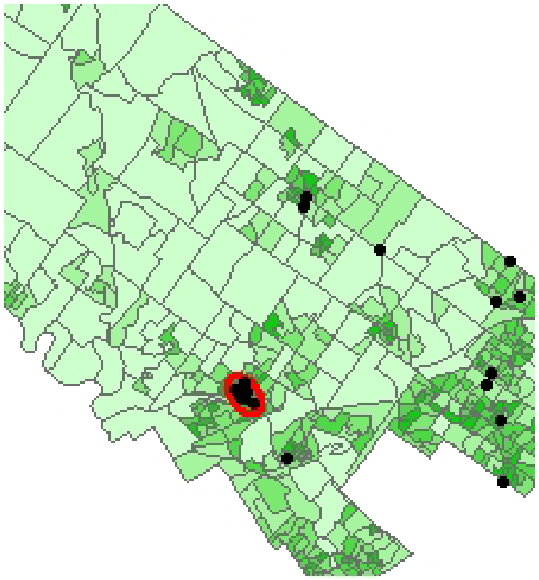
NNHC
(P-value = .01, Count >=5)



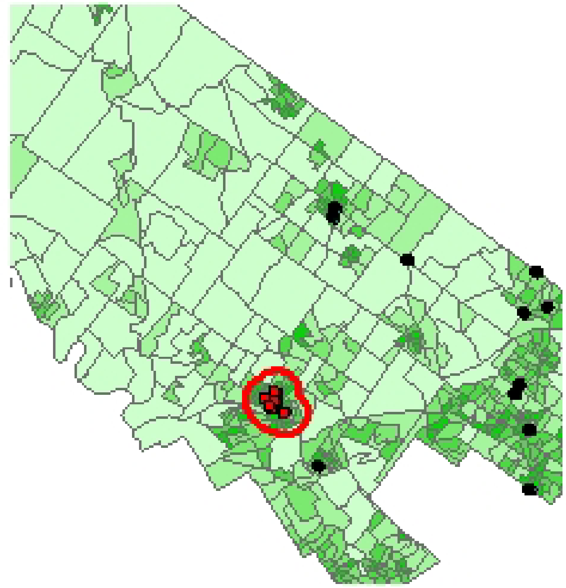
K-Function
(One-Half Mile radius, Count >=5)

**Montgomery County AFSP Clusters**

P-Value <= .01

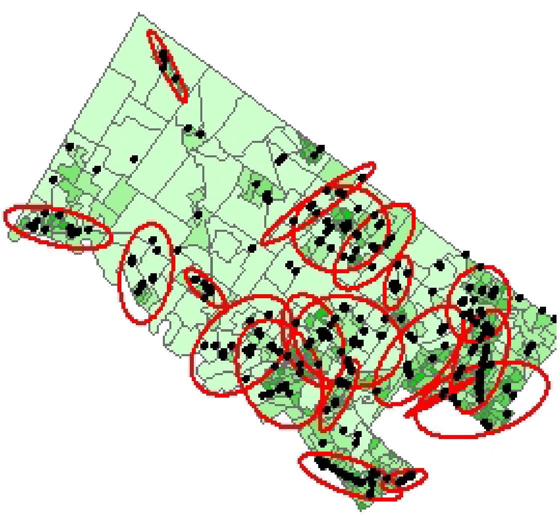


NNHC
(P-value = .01, Count >=5)

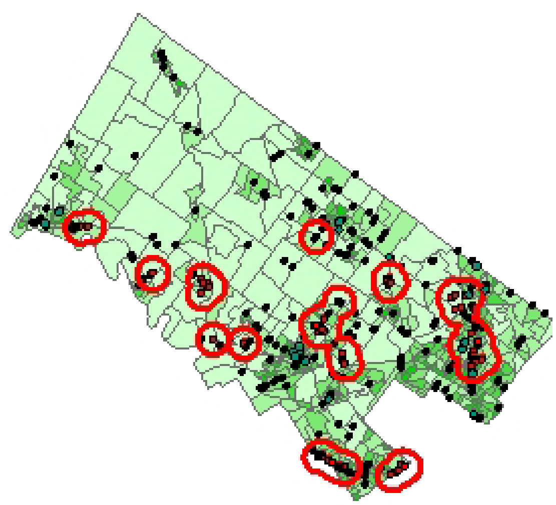


K-Function
(One-Mile radius, Count >=5)

**Montgomery County Bank Clusters**

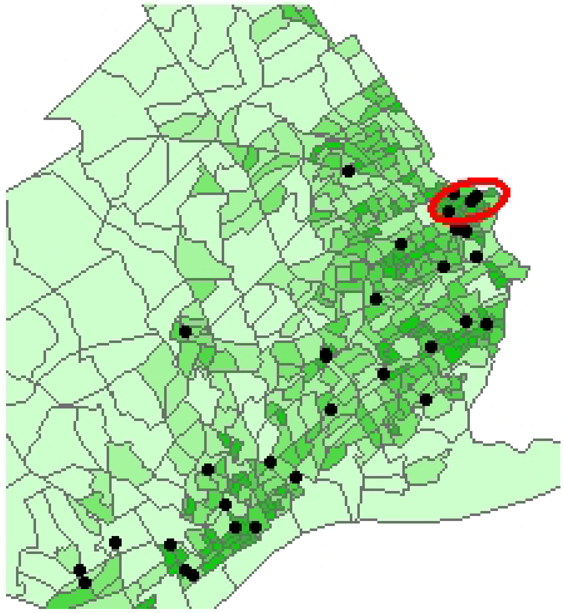


NNHC
(P-value = .01, Count >=5)

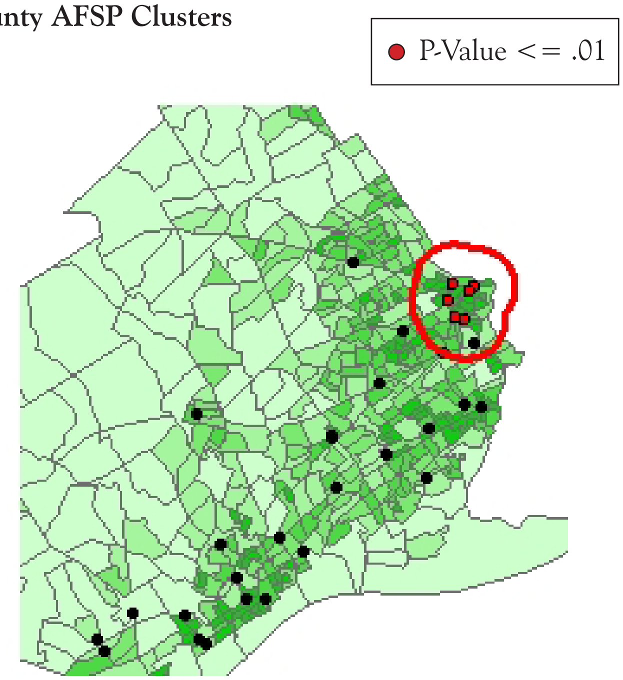


K-Function
(One-Mile radius, Count >=5)

**Delaware County AFSP Clusters**

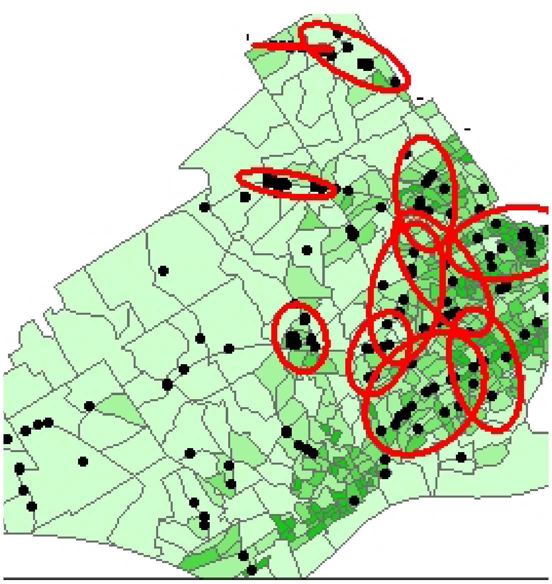


NNHC
(P-value = .01, Count >=5)

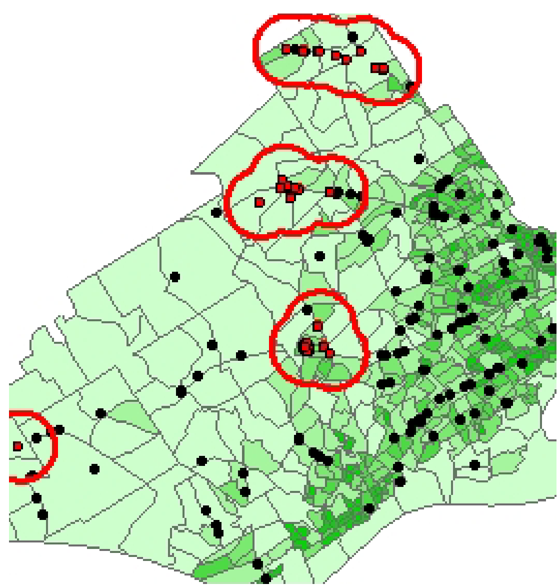



K-Function
(One-Mile radius, Count >=5)

**Delaware County Bank Clusters**



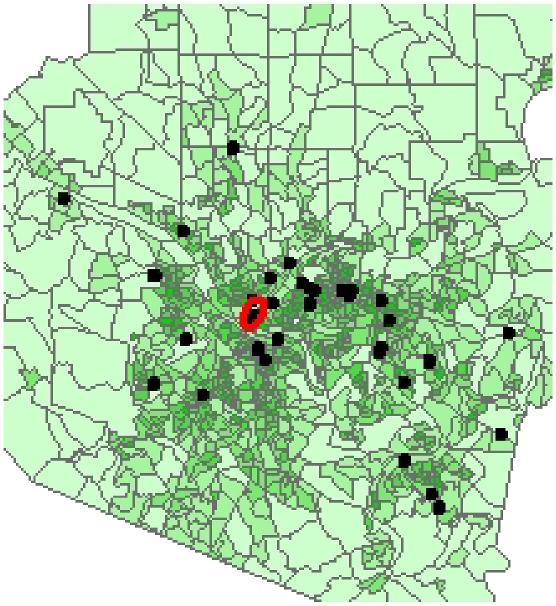
NNHC
(P-value = .01, Count >=5)



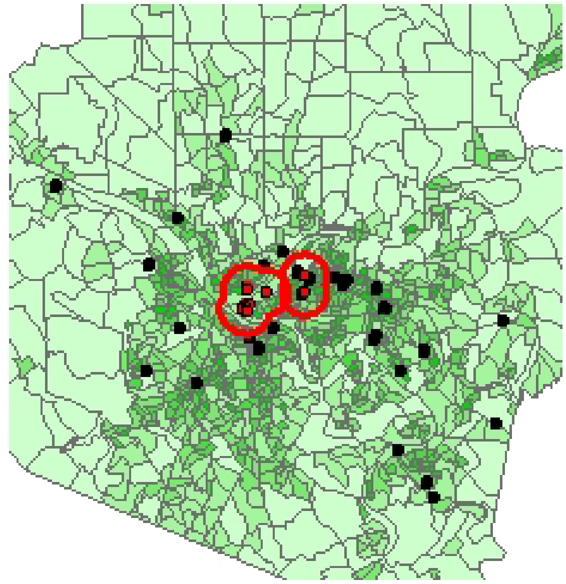
K-Function
(One-Mile radius, Count >=5)

**Allegheny County AFSP Clusters**

● P-Value <= .01

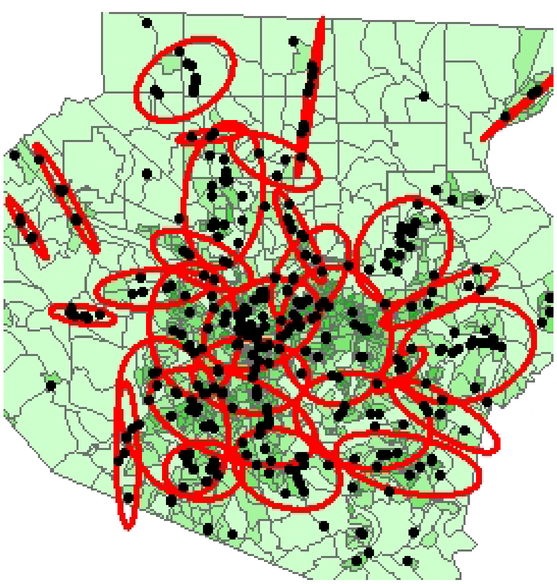


NNHC
(P-value = .01, Count >=5)

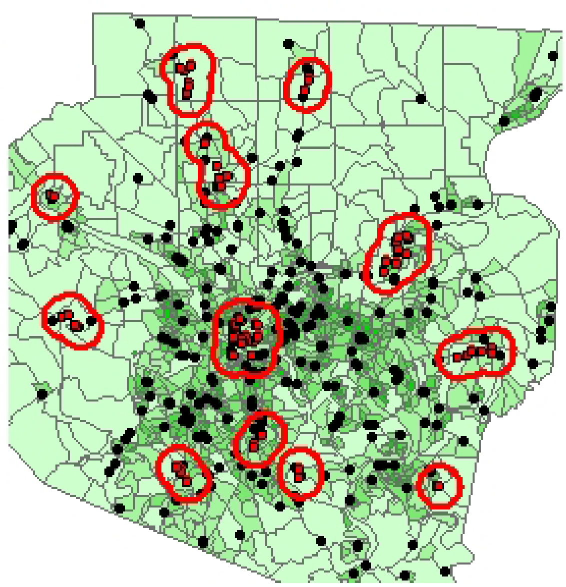


K-Function
(One-Mile radius, Count >=5)

**Allegheny County Bank Clusters**



NNHC
(P-value = .01, Count >=5)



K-Function
(One-Mile radius, Count >=5)

**Overlapping of AFSP and Bank Clusters**

As shown in the maps on the previous pages, there is a lack of overlapping of AFSP and bank clusters in the more suburban counties, using the K-function approach. Montgomery County contains bank clusters primarily in its western portion, with an AFSP cluster in the Norristown area. Delaware County exhibits an AFSP cluster in the Upper Darby area close to the county border with Philadelphia, while its bank clusters are found in the center, northern, and western portion of the county. Overlapping of AFSP and bank clusters does occur in Philadelphia and Allegheny counties. For Philadelphia County, the overlapping is found in Center City Philadelphia, while the overlapping in Allegheny County is in downtown Pittsburgh.[55]

[55] Refer to the text for a discussion of AFSP and bank clusters overlapping in centers of business activity.