This is a manuscript of an article from Rongerude, J., Haddad, M., 2016, "Cores and Peripheries: Spatial Analysis of Housing Choice Voucher Distribution in the San Francisco Bay Area Region, 2000–2010". *Journal of Housing Policy Debate* 26(3); 417-436. http://dx.doi.org/10.1080/10511482.2015.1128958 . Posted with permission.

# Cores and Peripheries: Spatial Analysis of HCV Voucher Distribution in the San Francisco Bay Area Region, 2000 - 2010

Jane Rongerude (Corresponding author)
Department of Community and Regional Planning
Iowa State University

Mônica A. Haddad
Department of Community and Regional Planning
Iowa State University

## Introduction

The Housing Choice Voucher (HCV) program is touted as a mechanism for providing low income households with mobility, and through that mobility, access to higher opportunity neighborhoods (HUD, 2006; McClure, 2014; Winnick, 1995). However, the success of the voucher program requires an existing supply of available housing units with landlords who are willing to participate in the program. Tight housing markets, where households are most in need of rent subsidies, are typically also the most difficult markets to find available units and willing landlords, resulting in limited choices for HCV households (Basolo and Nguyen, 2005; Khadduri, 2005). In these conditions, we can expect to find that voucher holders behave like other housing seekers and search out locations where they can maximize their housing dollars in terms of both unit characteristics and locational amenities.

The San Francisco Bay Area region is known for high land values and housing markets where not just low income households, but middle class households, struggle to find housing they can afford. The Harvard Joint Center for Housing Studies has created an interactive map that identifies the percentage of renter households in every CSMA in the country that are cost burdened, in other words, that pay more than 30% of their household income on housing.[i] For the Bay Area region, the numbers are universally high, ranging from 44.3% in the Napa Metro Area to 58.5% in the Stockton-Lodi Metro Area. Given the high level of competition for affordable units, one can assume that voucher holders in the Bay Area face struggle to find units in any condition in almost any neighborhood.

In contrast to other metropolitan regions where concentrated poverty is primarily located in the urban core or inner ring suburbs, the pattern of wealth and poverty in the Bay Area is more fragmented, and the location of poor places does not follow an easily identifiable pattern. High poverty areas may be located in the urban core neighborhoods of the region's largest cities, in unincorporated communities just outside city boundaries, or in edge cities located on the region's periphery. These places may be spatially isolated or located in close proximity to much wealthier places. In this context, it is not enough to analyze HCV distribution merely in terms of moves from the urban core to suburban communities. Opportunity and poverty exist in both places. Instead, this paper investigates whether the communities where HCV households locate share a particular set of characteristics, regardless of their location within the larger region.

There is clear evidence in the literature that HCV households are not randomly distributed in space, but tend to cluster, making space a critical factor in understanding HCV distributions (Park, 2013; Patterson and Yoo, 2012; Song and Keeling, 2010; Wang and Varady, 2005; Wang, Varady, and Wang, 2008). The first part of this study utilizes spatial analysis to investigate the distribution of voucher holders across the region. Unlike previous studies that used a technique called hotspot analyses to identify instances of statistically significant clustering within a given area (Song and Keeling, 2010; Wang and Varady, 2005; Wang et al, 2008), this study compares locational changes over time using kriging and kernel density. Kriging recognizes areas where points closer together are more likely to share a given characteristic than points closer apart. Both approaches make it possible to compare spatial distributions of HCV households across a large area, even when the geometries of the data are different for each year.

This study introduces housing researchers and policy makers to a methodological approach that addresses one common challenge related to incompatible data that researchers face when working with spatial analysis. This challenge is known in the field of geostatistics as a *change of support problem* (Gotway and Young, 2002) and can occur in various ways. In this study the challenge occurs when one uses a spatial variable that is available for two different years with a similar spatial scale, but with a different geometry for each year. For example, points representing individual HCV holders that changed locations between 2000 and 2010, and block groups with boundaries that also changed between 2000 and 2010. Our methodological approach also makes it possible to consider spatial dependence among variables in the regression analyses.

The second part of this study investigates the characteristics of the communities where voucher holder were located in 2010 and tests for a set of shared characteristics related to race, income, poverty, and housing prices. This is intended to illuminate whether voucher holders are accessing higher opportunity communities within the region. Similar to an earlier study by Park (2013), our methodological approach uses spatial regressions and spatial statistics to, “shed light on the impact of factors affecting voucher locational outcomes (p. 452)” and to control for spatial dependence.

Our findings are twofold. First, we found that from 2000 to 2010, the uneven distribution of voucher holders across the region continued and that the number of HCV households in certain areas increased at higher rates than others. Furthermore, areas with high densities of voucher holders could not be categorized as a single type of community: these concentrations occurred in urban core neighborhoods and inner ring suburbs, but also in communities on the regional periphery. Next we found that despite the locational variation of these places, the communities where voucher holders were living still shared a common set of descriptive characteristics. Voucher holders were more likely to locate in areas characterized by lower housing prices, lower percentages of educated people, higher rates of poverty[ii], and higher percentages of African American households when compared to the region as a whole. These findings suggest that voucher holders are moving to low opportunity areas, and confirm the findings of previous HCV studies: mobility is not enough. In tight housing markets with fragmented geographies of poverty and opportunity where low income communities are often already located in close physical proximity to more affluent communities, housing vouchers alone cannot provide low income households with access to opportunity. This may result from the lack of affordable housing in higher opportunity neighborhoods, and thus, policy makers also should address the issue of affordable housing supply.

## HCV locational outcomes: A review of the literature

HUD describes the HCV program as, "…the federal government's major program for assisting very low-income families, the elderly, and the disabled to afford decent, safe, and sanitary housing in the private market (HUD)." Since it first began as Section 8 in 1983, the program has provided millions of low income households with access to housing that is safe, sanitary, and affordable, and from this perspective the program has been highly successful (Varady, 2010). The vast majority of households who receive vouchers are able to secure housing where they pay less than 30% of their household income for rent and even more are able to secure housing where they pay less than 40% (McClure, 2014).

More recently, the program has been touted for its potential to provide low income households with access to higher opportunity neighborhoods, and as a result, decrease concentrated poverty and racial segregation in low income communities. Embedded within these policy objectives is a set of normative assumptions about how voucher holders will behave and what the market can accomplish. It assumes that voucher holder privilege neighborhood affluence above other factors in their housing search, that there are available units in these communities that meet the needs of HCV housing seekers, and that the market is free from discrimination and other barriers that would prevent households from accessing their desired units. These assumptions have proven to be problematic. "[T]he free market assumptions inherent in the voucher program are not always met in reality: minority voucher holders rarely escape poor, segregated communities (DeLuca, Garboden, and Rosenblatt, 2013, p. 269)."

Research into the locational distributions of HCV households reveals that in practice deconcentration outcomes have fallen short of expectations.[iii] In 2003, the first national study of HCV household distribution found that HCV participants as a whole were living in lower poverty neighborhoods than households living in fixed public housing units. However, they were still slightly more likely to live in the urban core than in suburban neighborhoods (Devine, Gray, Rubin, and Taghavi, 2003). A recent update to this study again concluded that the program does not result in significant outmigration from high poverty neighborhoods into low poverty neighborhoods (McClure, 2014). In other words, there is no evidence at the national scale that the HCV program has contributed to a significant redistribution of poverty.

What about a smaller scale? Are results different? In Chicago, Oakley and Burchfield (2009) found that HCV recipients tended to cluster in neighborhoods not significantly different from the public housing communities they initially occupied. Voucher housing in the Chicago area, like public housing, was predominantly clustered in census tracts characterized by high concentrated poverty and a high percentage of African-American residents. Other research has confirmed concentrations of HCV households in high poverty, racially segregated urban communities in places such as: Cleveland, New York, Cincinnati, Baltimore, (Park, 2012; Wang et al, 2008) and Jefferson County, Kentucky (Song and Keeling, 2010). At least one study has identified an instance where households were able to move into less racially segregated areas. In Patterson and Yoo's study of upstate New York, reclustering outside the city occurred only part of the time (2012). While noteworthy, this instance appears to be the exception in the literature.

Given the enduring legacy of racial discrimination in US housing markets, race and ethnicity play significant roles in determining outcomes for HCV households (DeLuca et at, 2013; McClure, 2014; Patterson, 2011). Voucher holders tend to remain clustered in poor and racially segregated areas even when they move outside the urban core. In communities where assisted households are mostly Black and other residents are mostly white, HCV households are much more likely to live in distressed neighborhoods (Pendall, 2000). Although every racial and ethnic group has households living in low-poverty neighborhoods, Black and Hispanic families are more likely than white families to live in high poverty neighborhoods. Alternatively, white families are more likely to live in low-poverty neighborhoods (Basolo and Nguyen, 2005; Devine, Gray, Rubin, and Taghavi, 2003). In the Patterson and Yoo study (2012), the majority of the African American voucher holders were in fact moving out of segregated communities of color into historically white neighborhoods; however, some clustering continued to occur, suggesting that re-location was taking place but not integration into the new communities (2012).

The HCV program provides the means for low income households to affordably rent units in the lower 50% of their local housing market. Although voucher holders can be found in more than 80% of all urban census tracts with existing rental units that meet FMR requirements (Devine, Gray, Rubin, and Taghavi, 2003), these affordable units are not evenly distributed within these census tracts or across space. Within major metropolitan areas, only 30% of the affordable units are located in low poverty census tracts (McClure, 2008). In an early study of HCV distributions, Pendall (2000) found that HCV users concentrated in distressed neighborhoods when rental housing is concentrated there, although overall, they avoid distressed neighborhoods with very low rents. The lack of available and affordable units in high opportunity locations suggests that the distribution of HCV households cannot be fully explained by personal choice or individual barriers. As voucher holders search for available units, supply matters.

Like all renters, voucher holders need available units and units they can afford. In their analysis of HCV household distribution in Santa Ana, California, Basolo and Nguyen (2005) consider the relationship between the price of rent and the neighborhood conditions where voucher holders relocate. They find that when voucher holders lived in better neighborhoods, it was not because of their mobility, but because they were living in neighborhoods that charged higher rents. Furthermore, HCV households in their study identified a lack of available houses to rent as their primary obstacle to mobility. To improve their neighborhood conditions, voucher holders had to be willing to pay higher rents and had to find an available unit. Given the low vacancy rates in Santa Ana, these two conditions were not always possible to meet.

In 2005 Wang and Varady introduced the use of hotspot analysis to identify clustering of HCV households. In a study of the Cincinnati area, they identified clusters of HCV households in three types of communities, including a suburban census tract with low poverty rates. However, further investigation of that census tract revealed that the cluster corresponded with a Low Income Housing Tax Credit development, calling attention to the need for including context in studies analyzing the locational choices of HCV households (see also Varady, Wang, Wang, and Duhaney, 2010; Wang et al., 2008). Song and Keeling's study in Jefferson County, Kentucky, used spatial analysis to demonstrate that even in smaller housing markets, HCV households continue to concentrate in the city core in African American neighborhoods with high poverty rates (2010). Finally, in a study of Cleveland, Park used spatial analysis to investigate the relationship between the clustering of voucher households and the racial and economic conditions of the places where they clustered (2013). He expanded on earlier spatial methodologies by incorporating Geographically Weighted Regressions (GWR) into the analysis. Park concluded not only that HCV households cluster, but that they are more likely to cluster in communities with higher percentages of African Americans, higher percentages of available affordable units, and better access to transportation.

## Study context

Typically, the San Francisco Bay Area is defined as a nine-county area encompassing four different US Census Metropolitan Statistical Areas (MSAs). More recently, academic research on the region has begun to move in the direction of mega-regional analysis (Brenner and Pastor, 2011; Innes, Booher, and Di Vittorio 2010; Schafran, 2014). Given the high cost of living in many of the region's urban nodes and the increases in poverty that have been identified at the exurban edge of the region and in particular the counties just east of the traditional Bay Area region (Brookings, 2011; Soursourian, 2012), this study anticipates that voucher holders may be moving beyond the traditional boundaries of the region in their search for housing. As a result, we use a fourteen county mega-region that includes the traditional nine county area plus an additional five counties with a known Bay Area outmigration (Schafran, 2014). The final study area encompasses the following 14 counties: Alameda, Contra Costa, Marin, Merced, Napa, Sacramento, San Francisco, San Joaquin, San Mateo, Santa Clara, Solano, Sonoma, Stanislaus, and Yolo. Figure 1 displays the study area with the traditional core counties and the peripheral counties differentiated in the legend.

[FIGURE 1 ABOUT HERE]

From 2000 – 2010, this expanded Bay Area region had an average growth rate of 8.82% (US Census). The growth rates within the region were lowest for San Mateo (1.60%), Marin (2.07%), and San Francisco (3.67%) counties. They were highest in the San Joaquin, (21.59%), Merced (21.49%), and Yolo (19.09%) counties. While the region continued to grow in terms of population and physical area, it also grew in terms of poverty and inequality. The counties with the highest growth rates also reported the highest poverty rates in the region. In San Joaquin County, 16% of individuals in the county were living below the poverty line in 2010. In Merced County that number was 21.8 % and in Yolo County it was 17.1%. The lowest poverty rates in the region were 7% in Marin and San Mateo Counties (ACS). Increased poverty rates in growing counties means these areas are experiencing significant increases in the total number of people living in poverty. See Table 1 for a comparison of poverty and growth rates for each county in the study area.

[TABLE 1 ABOUT HERE]

According to a study by the Federal Reserve Bank of San Francisco, the Bay Area experienced both an increase in poverty and an increase is suburban poverty from 2000 to 2009 (Soursourian, 2012).[iv] Where the region's urban cores lost five high poverty census tracks between 2000 and 2009, the suburbs gained two (Brookings, 2011). The region's total household poverty rate increased 1.1 percentage points and poverty rates increased across almost all groups. Changes in poverty levels also varied by racial and ethnic group and by nativity status. Asians and the foreign- born population living in suburban areas did not experience an increase in poverty levels, but the poverty rates for suburban Blacks and urban Hispanics each rose more than two percentage points. Changes in poverty levels also varied across space. The number of people living in poverty rose 16 percent in the suburbs, compared to 7 percent in urban areas. Blacks and Hispanics saw the greatest percentage growth in suburban poverty. The movement of poor households from urban areas to the suburbs, a process known as the suburbanization of poverty, was greatest for the region's poor Black population. The share of the poor Black population living in the suburbs increased more than 7 percentage points. The next highest group, Asians, increased 2 percentage points (Soursourian, 2012).

**The spatial distribution of HCV households**

This study was designed to identify the changing spatial distributions of HCV households over time and the characteristics of the communities where these households locate. This required not just data about HCV households, but also socioeconomic data for our study area. We used block group level Census data rather than tract level data in order to attain a more fine-grained understanding of the places where HCV households locate (Oakley and Burchfield, 2009; Park, 2013) and the boundaries of those places (Talen and Koschinsky, 2014). For the 2000 and 2010 HCV locational data, we used administrative records provided by HUD and pulled from HUD's PIH Information Center (PIC) data system.[v] The HUD dataset identified each HCV household within our study area by latitude and longitude. We then transformed those coordinates into points through a geocoding process. To facilitate meaningful comparisons between the two years, we normalized the HCV data by calculating the annual percentage of HCV households per block group, creating the variable: 'Percentage of HCV households.'[vi] We did this using the spatial join tool available in ArcGIS software.

After the geocoding process, the final 2000 point shapefile had 58,619 HCV households and the final 2010 point shapefile had 99,212 HCV households. The total number of vouchers in our dataset increased by approximately 70 percent between 2000 and 2010. Some of this difference can be accounted for by an expansion of the voucher program during this period, including some conversion of public housing units into vouchers through HOPE VI. However, the growth is mostly accounted for by improvements in reporting by local public housing authorities (PHAs). The annual reporting is voluntary and not every PHA fully participates. However, by 2010 every Bay Area PHA was reporting to HUD using the PIC system and could be considered compliant.

When working with multiple datasets in a study area of this size, differences in scale and/or geometry can become an issue. As Le Gallo and Chasco (2012) explain, “in empirical studies dealing with spatial data, researchers are frequently confronted with data available at different spatial scales (p. 281).” Confronting data with different scales and geometries within a single bounded area is an example of a change of support problem. Change of support problems are related to the modifiable areal unit problem (MAUP). MAUP occurs when “data aggregated to a particular set of geographical regions changes if the same data are aggregated to a different set of geographical regions” (Peeters and Chasco 2006 p. 259), and aggregation typically decreases variability among units. As described by Gotway and Young (2002), the MAUP refers to two interrelated problems. First, there is the aggregation effect, meaning that when “the different inferences [are] obtained when the same set of data is grouped into increasingly larger areal units” (p. 633). Second, there is the zoning effect, which refers to when “variability [is observed] in results due to alternative formations of the areal units leading to differences in unit shape at the same or similar scales” (p. 633).

The change of support problem and the aggregation effect in this study are the result of point and polygon spatial variables that were available for both 2000 and 2010, but had different geometries for each year (i.e., the location and quantity of polygons and points for 2000 did not match the location and quantity of polygons and points for 2010). Figure 2 illustrates the change of support problem and the aggregation effect in this study.

[FIGURE 2 ABOUT HERE]

Le Gallo and Chasco (2012) call attention to two problems that could arise when analyzing variables from different spatial levels at one single level. The first one is a statistical problem: "if data was aggregated […] some information would be lost and the statistical analysis would lose power (p. 282)." The second one is a conceptual problem: "the interpretation of the results may be affected by the fallacy of the wrong spatial level, which consists of analyzing the data at one level, in order to draw conclusions at another level (p. 282)." Kriging can provide a solution when data are at different scales or have different geometries (Le Gallo and Chasco, 2012) because it changes the support of a variable by creating a new variable (Gotway and Young, 2002). Calvo and Escolar (2003), also suggest that Geographically Weighted Regressions (GWRs), provide a possible solution to these problems of spatial aggregation bias; however Páez, Farber and Wheeler (2011) indicate that the GWR approach can cause spurious correlations.

Kriging "is unique among the interpolation methods in that it provides an easy method for characterizing the variance, or the precision, of predictions (ESRI)," and it has been applied to numerous spatial problems that analyze data at different scales and with different geometries. With this technique, "the surrounding measured values are weighted to derive a predicted value for an unmeasured location. Weights are based on the distance between the measured points, the prediction locations, and the overall spatial arrangement among the measured points (ESRI)."

Kriging has been used most frequently to address biophysical phenomena; however, more recently researchers have applied this approach to other fields including housing prices (Ahlfeldt, 2011); land value (Tsutsumi and Seya, 2008; Wendland, 2015); noise (Chasco and Le Gallo, 2013); and hedonic modeling, in particular those addressing air pollution point data and housing locations (see Anselin and Lozano-Gracia, 2008; Le Gallo and Chasco, 2012; Mínguez, Montero, and Fernández-Avilés, 2013). In the policy realm, kriging was used to create an environmental index for the City of Madrid (Montero, Chasco and Larraz, 2010). We opted to use kriging to address the change of support problems in this study.[vii]

### *Kriging*

Some spatial interpolation techniques assume that data is spatially correlated. As a result, we had to establish that the 'Percentage of HCV households' variable displayed clustering patterns before beginning the spatial analysis. Establishing spatial autocorrelation makes it possible to assume that locations closer in distance are closer in value than locations farther apart, in other words, that clustering is occurring (ESRI, 2013a). We tested the 'Percentage of HCV households' variable for spatial autocorrelation in 2000 and 2010 using the Moran's $I$ and the Geary's $c$ tests.[viii] We used two spatial matrices in these tests: a simple binary queen contiguity matrix and a k-nearest-neighbors matrix.[ix] The results shown in Table 2 were computed using 999 permutations, and the p-value for all eight statistics was 0.001. Based on these results, we could confirm a positive spatial autocorrelation in the distribution of 'Percentage of HCV households' in the San Francisco Bay Area for both years. The fact that Moran's $I$ and the Geary's $c$ presented very similar results confirmed that these results were robust.

[TABLE 2 ABOUT HERE]

Once we established spatial autocorrelation, we could begin the process of spatial interpolation. Using the 2000 and 2010 points representing the annual ‘Percentage of HCV households’ per block group, we estimated surfaces using a kriging spatial interpolator.[x,xi] We estimated six kriging surfaces, applying a second-order polynomial to remove the global trends in the data for all six.[xii] We used three variogram models for each year: spherical, Gaussian and quadratic.[xiii] Table 3 summarizes the models’ parameters. Identifying the best model was not straightforward.[xiv]. We opted to use the ordinary quadratic model for 2000 and the ordinary Gaussian model for 2010 because both produced the best overall results.

[TABLE 3 ABOUT HERE]

The next step was to create the surfaces for analysis. Kriging creates surfaces, enabling us to compare 2000 and 2010 despite the different geometry and quantity of the polygons from each year.[xv] Figure 3 shows the two kriging surfaces: 2000 and 2010.[xvi, xvii] When visually comparing the two surfaces, a number of differences stand out. The region overall is much darker in 2010, the result of the overall increase of vouchers in our dataset. At the same time, the rate of change is not uniform. The highest value areas (darkest color) cover more area in 2010 and are often surrounded by expanded medium dark areas. These areas emerge in urban job centers and in inner ring suburban locations surrounding the Bay; however, they also emerge in the agricultural areas in the northern part of the region and in the counties along the region’s eastern periphery.

[FIGURE 3 ABOUT HERE]

Now that we had surfaces for each year, the next step was to analyze for changes over time. Using the ArcGIS 10.2 Map Algebra tool, we subtracted the 2000 surface from the 2010 surface and created a new, final surface (see Figure 4). The symbology depicted in Figure 3 is based on the average difference of values: ‘low increase’ represents values 90 percent smaller than the regional average difference, ‘medium increase’ represents values between 90 and 110 percent of the regional average difference, and ‘high increase’ represents values higher than 110 percent of the regional average difference.

[FIGURE 4 ABOUT HERE]

Figure 4 confirms that the ‘Percentage of HCV households increased across most of the region. The concentrations of ‘percentage of HCV households’ increased in many of the region’s urban job centers and in inner ring suburbs surrounding the Bay. Again, increased concentrations are visible around San Francisco, and the East Bay job centers. However, they also increased in many communities located in the region’s periphery, and in particular the eastern edges of the region from Yolo County near the northeastern edge of the region down to Merced County, the southern most county on the region’s periphery. This method revealed that the density of HCV households increased in both the cores and the peripheries of the region. Our next step is to confirm these results using kernel density estimation.

### ***Kernel density estimation***

Kernel density estimation is another spatial interpolation technique that identifies areas with greater concentrations of a phenomenon under examination.[xviii, xix] It is also a powerful tool for geovisualization. This density mapping enables us to “expresses the distribution of point values over a surface without limiting the analysis to census tracts or other geo-political boundaries (Mellow, Schlager, and Caplan, p. 419, 2009).” With this approach, a circle is centered at each cell of the study area and then a HCV household’s density value is calculated for each cell. It is important to point out that the values from the figures below are not comparable to the values in the preceding figures. In other words, kernel density accomplishes something very different than kriging. Where kriging uses the ‘percentage of HCV households’ variable in the analysis, kernel density uses each individual HCV household and looks for the density of points (and their attributes) in the region around each cell. Using individual points for kernel density is optimal for this study because it allows us to assess the “rawest measure of local concentration (Metraux et al, p. 245, 2007)” in addition to the percentage measurement. The values, then, are a reflection of the density of individual HCV households.

Despite the similarity between kriging and kernel density estimation, “the intention and the outcomes are quite different (O’Sullivan and Unwin, 2010, p. 234).” The intention in kriging is to predict unmeasured values while kernel density is a visualization tool that measures intensity. This study uses kriging to allow a comparison between 2000 and 2010 HCV data, based on the predicted surfaces. Kernel density estimation, a more established tool within housing policy research, is then used to show changes of HCV household intensity across the region. With regards to outcome, additional points included in kernel density estimation do not affect the weights associated with each event. In kriging, the inclusion of additional points reduces the weight of each observation; and as a result, kriging is more effective in dealing with spatial variability. We have included both methods in our analysis of HCV distribution in order to ensure rigor in our study and to investigate the potential for kriging as a tool for housing research.

The 2000 and 2010 kernel density surfaces are displayed in Figure 5.[xx] Again, significant differences are observable between the two surfaces. Most notably, the darkening in San Francisco, Silicon Valley, and the East Bay employment corridor as well as the Sacramento region in the northeast and urban centers running from north to south within the peripheral counties to the east of the region. We again observe the pattern of voucher growth in both the urban cores and the more traditionally rural peripheries of the region.

[FIGURE 5 ABOUT HERE]

Using the same Map Algebra procedure described earlier, we subtracted the 2000 kernel density surface from the 2010 surface. Figure 6 displays the final map. The East Bay and Silicon Valley employment corridors show significant increases in HCV households as expected, as does the Sacramento metropolitan area. Furthermore, while the increases in peripheral areas along the eastern edge of the study area are pronounced in Figure 6, increases in the northwestern agricultural areas of the region are also observable in this figure. Figure 4 displays a finer grain of variability than Figure 6, however, the areas of high concentration remain the same in both. This confirms our findings from the kriging analysis: although the concentration of 'individual HCV households' increased in several communities located along the region's periphery, it also increased significantly in the region's urban job centers. The second part of this paper will consider additional characteristics of the receiving communities where HCV households locate.

[FIGURE 6 ABOUT HERE]

## Community characteristics

This phase of our analysis uses spatial regression models to test for common characteristics among the communities with high values of HCV households. While regression models are used widely in housing studies, spatial regressions have a more limited presence. They are frequently used in housing studies that estimate hedonic models, but are relatively new in the context of HCV mobility. This analysis again utilizes block group level data from the 2000 Census as well as the percentage of HCV households at the block group level. Table 4 provides an overview of the variables we used in our regressions.

[TABLE 4 ABOUT HERE]

The dependent variable, 'PER_HCV,' is the 2010 percentage of HCV households. The independent variables related to housing values and socioeconomic characteristics are from the 2000 Census: percent total population with income in 1999 below poverty level; percent total population (all races) Black or African American alone or in combination; percentage population 25 years and older with graduate degree (master's professional or doctorate); and median value (dollars) for all owner-occupied housing units. We used a ten-year time lag between the dependent and independent variables to minimize any possible problem of endogeneity.[xxi] Table 5 depicts the Pearson Coefficient matrix of all variables included in the regressions. All correlations have the expected signs, and the strongest positive correlation is between median housing price and level of education.

Our independent variables were informed by a literature review of empirical studies that used regression models for understanding low-income housing mobility within a given metropolitan area (Basolo and Ngyuen, 2005; Faber, 2013; Meltzer, 2013; Park, 2013; Fitzgerald and Vitello, 2014). Variables related to race, education, and income level were consistently included in these studies. Initially we also included percent family household heads that are female (includes only family households) and rental housing vacancy rate (number of vacant for rent dwellings divided by sum of occupied rentals and vacant for rent dwellings) as independent variables. However, their inclusion caused a multicolinearity problem, i.e., there was a high correlation between two or more independent variables, making it difficult to identify the relationship between independent and dependent variables.

[TABLE 5 ABOUT HERE]

### *Spatial regressions*

Because the dependent variable, 'PER_HCV,' presented spatial autocorrelation in its distribution, we needed to identify whether a spatial dimension was also necessary for the regression models.[xxii] As Ertur, Le Gallo and Baumont (2006) explain, "it is well known that the presence of spatial dependence […] leads at best to unreliable statistical inference based on ordinary least squares (OLS) estimations (Ertur, Le Gallo, and Baumont, 2006, p. 5)." By identifying the existence of spatial dependency, we can avoid these potential errors. We used the ordinary least squares (OLS) model with four independent variables for this test. We estimated the OLS model, as follows:

$$PER\_HCV_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_4 x_{4i} + \varepsilon_1 \; i = 1, 2, \ldots 6{,}124 \qquad (1)$$

where $PER\_HCV_i$ is the percentage of HCV households for each census block group i, $x_{1i}$ to $x_{4i}$ are the independent variables representing percentage of African Americans, percentage with graduate degrees, percentage in poverty, and median housing price; $\beta_0$ to $\beta_4$ are the unknown parameters to be estimated; and $\varepsilon_1$ represents the error terms.

When estimating the OLS, we tested for spatial dependence in the model specification using Lagrange Multiplier tests (LM).[xxiii] Depending on the results of these tests, it may be necessary to estimate the spatial lag and/or spatial error models. A spatial lag model "incorporates a spatially lagged dependent variable on the right hand side of the regression model (Anselin, 2002, p. 247)." As a result, "observed outcomes are simultaneously determined with outcomes for neighboring areas (Mobley, Root, Anselin, Lozano-Gracia, & Koschinsky, 2006, p. 13)." In the case of our study, this means that a block group with a high percentage of HCV households would affect the nearby block groups by elevating their percentage of HCV households (Bauer et al, 2013). A spatial error model indicates that, "spatial error autocorrelation is either modeled directly, following the general principles of geostatistics, or by utilizing a spatial autoregressive process for the error term (Anselin, 2002, p. 248)." Moreover, "in a spatial error model, unobserved factors in neighboring areas are correlated leading to correlation in the error term across space (Mobley et al, 2006, p. 13).” In other words, there would be missing independent variables in the multivariate analysis in Equation 1, and these variables would, “follow a meaningful spatial pattern (Páez and Scott, 2004, p.55).”

A spatial lag model requires an additional explanatory variable in the specification. For this study, it is the spatially lagged percentage of HCV households. The model, that is estimated using maximum likelihood, is written as:

$$PER\_HCV_i = \beta_0 + \rho_1 wPER\_HCV_{1i} + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_4 x_{4i} + \varepsilon_2 \quad (2)$$

$$i = 1, 2, \ldots 6{,}124$$

where all elements are defined as previously, and $PER\_HCV_i$ is the set of the spatially lagged dependent variable; $\rho$ is the spatial autoregressive parameter to be estimated; and $\varepsilon_2$ represents the error terms.

Second, in the error model, the spatial dependence is in the error term $v$. It is also estimated using maximum likelihood. The equation follows:

$$PER\_HCV_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_4 x_{4i} + v_1 \quad (3)$$

$$v_1 = \lambda * wv + \varepsilon_3 \quad i = 1, 2, \ldots 6{,}124$$

where all elements are defined as previously; $\lambda$ is the scalar parameter expressing the intensity of spatial autocorrelation between regression coefficients; and $v_1$ and $\varepsilon_3$ represent the error terms.

We did all calculations in OpenGeoda using both the queen and six-nearest-neighbors spatial matrices, which allowed us to confirm the robustness of our results.

*Results*

Table 6 displays the estimation results of the queen spatial matrix. It does not include the estimates based on the six-nearest-neighbors spatial matrix because the results were very similar. The OLS model is depicted in the first column. The coefficient for ‘median housing price’ and ‘percentage with graduate degree’ were negative and significant, indicating that lower housing price and people with lower education level coincided with high ‘percentage of HCV households.’ The coefficients for ‘percentage in poverty’ and ‘percentage of African Americans’ were all positive and significant, indicating that higher shares of African Americans and people in poverty coincided with high ‘percentage of HCV households.’ Because the multicolinearity condition number is approximately 6, we can state that multicolinearity is not a problem in the model.

[TABLE 6 ABOUT HERE]

The adjusted $R^2$ showed that the OLS performed well with a value of approximately 15 percent. Next we examined the LM tests. Because the value for both the spatial lag model and the spatial error model were high and significant we decided to estimate both models, i.e., the spatial lag and the spatial error models.

The spatial lag model is displayed in the second column of Table 6. The spatial lag coefficient is positive and significant indicating that spatial autocorrelation between 'PER-HCV' levels in neighboring areas was not captured by the independent variables used in the OLS-based model. The coefficients for ‘median housing price’ and ‘percentage with graduate degree’ were negative and significant indicating that lower housing prices and people with lower educational attainment coincided with a higher ‘percentage of HCV householder.’ The coefficient for ‘percentage of poverty’ and ‘percentage of African Americans’ were positive and significant indicating that  higher percentages of people in poverty and of African Americans coincided with a high ‘percentage of HCV households.’ The spatial error model is displayed in the third column of Table 6. There is a positive and significant error spatial autocorrelation, and all the results are consistent with column 2, i.e., the spatial lag model.

Both spatial models provided an accurate description of the places where HCV voucher holders located in our study area; however, each provided slightly different insights into the possible explanations for this distribution. The lag model indicated that between 2000 and 2010, concentrations of HCV households existed in the SF Bay Area Region. Furthermore, a block group with a high percentage of HCV households affected the nearby block groups by elevating their percentage of HCV households. This supports existing studies that have found that HCV households often choose locations based on their existing social networks. The error model indicated that unobserved factors in neighboring block groups were correlated, resulting in correlation in the error term across the SF Bay Area Region. In other words, unknown factors were contributing to the outcome much more so than the concentration of vouchers. This paper has raised the question of whether housing prices affect the locational choices of HCV households and the error model strongly indicates that unobserved factors exist that should be examined in future analysis.

Based on the values of the log likelihood, the Akaike criterion, and the Schwartz criterion (see Table 6), we were able to establish that that the spatial error model provided the best fit of the three. This is because the log likelihood for this model was the highest and both the Akaike and Schwartz were the smallest (Anselin and Rey, 2014). This result was expected, as the robust LM tests indicated that the spatial error model seemed more appropriate for dealing with the spatial dependency in this study (a value of 77 against 3.39 for the spatial lag model).

Using these models, we were able to identify that areas with higher values of HCV households in 2010 had higher percentages of African Americans and of people in poverty, lower housing prices, and lower percentages of educated people when compared to the region as a whole in 2000. Poverty as a variable did prove to be significant. When combined with the three other variables that showed significance in this model, the analysis provides additional details about the lack of access to opportunity for people living in these areas. This part of the study identified a set of shared characteristics that hold true of all destination neighborhoods where voucher holders are locating across the region, bridging urban core, suburban, and exurban peripheral places.

**Discussion: Limits to mobility**

The HCV program provides low income households with access to the private market and enables individual mobility in the pursuit of affordable housing. However, the ability to access the market is not the same as the ability to access a housing unit. As Devine, Gray, Rubin, and Taghavi explain, “HCV rental assistance can only be provided where there is affordable housing (2003, p. 8).” This study finds that voucher holders are locating in communities with higher poverty rates and lower housing prices than the region as a whole. In tight housing markets like the San Francisco Bay Area, where competition is high for desirable units at almost any price point, mobility programs are not enough to provide HCV households with access to higher opportunity areas. This scenario underscores the need for policy makers to consider both household mobility and affordable housing supply within a given metropolitan context. Incentivizing affordable housing production increases choice for all low income households, whether they are HCV households desiring to move out of a higher poverty community or low income households desiring to remain in place but facing threats of displacement from the superheated housing market of the San Francisco Bay Area.

Varady and Kleinhans (2008) have argued that voucher holders will continue to concentrate in low opportunity, segregated communities until the distribution of affordable housing changes, and in fact, a recent study by Metzger found that the HCV program has perpetuated racial and economic segregation rather than relieved it (2014). The study presented in this paper demonstrated that voucher holders were not randomly distributed across the region, but that they occurred in higher values in communities with higher percentages of African American households, higher percentage of households living in poverty, and lower rates of educational attainment. The specific patterns of dispersal in the SF Bay Area reflects the documented suburbanization of poverty occurring during this period of time and the outmigration of African American households from the urban core to communities located in the regional peripheries (Schafran and Wegman, 2012). These findings call into question whether households that relocate outside the region's urban core neighborhoods are indeed improving their situations or instead following a larger trend of racial and economic redistribution within the region.

As a demand side rental subsidy, the HCV program is premised on the assumption that high amenity communities contain available units with rents at Fair Market Rental (FMR) rates. Researchers have identified that most neighborhoods in the country's 50 most populated MSAs contain some affordable units (Devine, Gray, Rubin, and Taghavi, 2003). However, we know very little about the location or availability of affordable units in affluent communities, despite knowing a great deal about the dispersal and density of HCV households. As long as the vast majority of available affordable units continue to be clustered in neighborhoods with limited opportunity—whether they are traditionally impoverished urban core neighborhoods or exurban communities where housing prices are lower, but access to transportation and other amenities is more limited—voucher holders will continue to cluster in these communities as well. Creating opportunities for low income households to move into higher opportunity communities requires changing the distribution of affordable housing units and further incentivizing the construction of affordable rental housing in desirable communities.

Finally, this study introduced a new methodology for investigating and analyzing the distributions of HCV households. Although we used this methodology to analyze the distribution of HCV households, a similar approach could be used to evaluate the distribution of affordable housing units, for example, in relation to the socioeconomic characteristics of a place. It also provides a tool for local PHAs or decision-makers interested in evaluating how HCV movements correlate with the availability and affordability of housing within their metropolitan region. In tight housing markets such as the SF Bay area, most HCV households will experience limited housing choice until the supply of affordable units is expanded. This type of spatial information about a metropolitan area can assist in the locational targeting of incentives for the construction of new affordable units.

## REFERENCES


Ahlfeldt, G. (2011). If Alonso was right: modeling accessibility and explaining the residential land gradient. *Journal of Regional Science. 51,* 318–338.

Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural Economics, 27, 247–267.* doi: 10.1111/j.1574-0862.2002.tb00120.x

Anselin, L. (1988). Spatial Econometrics: Methods and Models. Dordrecht, The Netherlands: Kluwer Academic.

Anselin, L. and Rey, S. (1991). Properties of tests for spatial dependence in linear regression models. *Geographical Analysis*, 23(2), pp. 112–131.

Anselin, L., & Le Gallo, J. (2006). Interpolation of air quality measures in hedonic house price models: spatial aspects. *Spatial Economic Analysis, 1*(1), 31-52.

Anselin, L., & Lozano-Gracia, N. (2008). Errors in variables and spatial effects in hedonic house price models of ambient air quality. *Empirical Economics*, *34*(1), 5-34.

Anselin, L. and Rey, S. (2014). Modern spatial econometrics in practice: A guide to GeoDa, GeoDaSpace and PySAL. *Paperback ed: GeoDa Press LLC*.

Basolo, V. & Nguyen, M. T. (2005). Does mobility matter? The neighborhood conditions of housing voucher holders by race and ethnicity. *Housing Policy Debate, 16,* 297-324. doi: 10.1080/10511482.2005.9521546.

Bauer, E. P., Arnio, A. N. & Wolff, K. T. (2013). Assessing the role of mortgage fraud, confluence, and spillover in the contemporary foreclosure crisis. *Housing Policy Debate*, *23,* 299-327. doi: 10.1080/10511482.2012.727843.

Benner, C., & Pastor, M. (2011). Moving on Up? Regions, megaregions, and the changing geography of social equity organizing. *Urban Affairs Review, 47,* 315-348. doi: 10.1177/1078087410391950.Behavioral Risk Factor Surveillance System BRFSS (2015) Retrieved March 03 2015 from http://www.cdc.gov/brfss/maps/faqs.htm

Chasco, C., & Gallo, J. L. (2013). The impact of objective and subjective measures of air quality and noise on house prices: a multilevel approach for downtown Madrid. *Economic Geography*, *89*(2), 127-148.

Dai , D., Zhang, Y., Lynch, C. A., Miller, T., and Shakir, M. (2013). Childhood drowning in Georgia: A geographic information system analysis. *Applied Geography, 37,* 11-22. doi:10.1016/j.apgeog.2012.10.006.

DeLuca, S., Garboden, P. M. E., and Rosenblass, P. (2012) Segregating Shelter: How Housing Policies Shape the Residential Locations of Low-Income Minority Families. *ANNALS of the American Academy of Political and Social Science*, *647,* 268-278. doi: 10.1177/0002716213479310.

Devine, D. J., Gray, R. W., Rubin, L., and Taghavi, L. B. (2003). Housing Choice Voucher Location Patterns: Implications for Participants and Neighborhood Welfare. U.S. Department of Housing and Urban Development Report. Accessed on 10/16/13 at: http://www.huduser.org/publications/pdf/Location_Paper.pdf.

Ertur, C., Le Gallo, J., & Baumont, C. (2006). The European regional convergence process, 1980-1995: Do spatial regimes and spatial dependence matter? *International Regional Science Review*, *29*(1), 3-34.

ESRI 2013 a. Retrieved October 20, 2013, from http://training.esri.com/Courses/GADataPatterns10_0/player.cfm?c=315

ESRI, 2013 b. Retrieved October 30, 2013, from http://training.esri.com/Courses/GAInterpolate10_0/player.cfm?c=341

ESRI 2013 c. Retrieved November 02, 2013, from http://training.esri.com/Courses/GAInterpolate10_0/player.cfm?c=341

ESRI 2015a Retrieved March 03, 2015, from http://resources.arcgis.com/en/help/main/10.2/index.html#//009z00000011000000

ESRI 2015b Retrieved March 10, 2015, from http://resources.arcgis.com/en/help/main/10.1/index.html#//0031000000m3000000

Faber, J. W. (2013). Racial dynamics of subprime mortgage lending at the peak. *Housing Policy Debate*, *23*(2), 328-349.

Feins, J. D. and Patterson, R. (2005). Geographic Mobility in the Housing Choice Voucher Program: A Study of Families Entering the Program, 1995–2002. *Cityscape: A Journal of Policy Development and Research, 8,* 21-47.

Fitzgerald, J., & Vitello, S. P. (2014). Impacts of the Community Reinvestment Act on Neighborhood Change and Gentrification. *Housing Policy Debate*, *24*(2), 446-466.

Galvez, M. (2010). "What Do We Know About Housing Choice Voucher Program Location Outcomes?: A Review of Recent Literature, " Urban Institute: Washington, DC. Retrieved December 15, 2015 from http://www.urban.org/UploadedPDF/412218-housing-choice-voucher.pdf.

Gotway, C. A., & Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, *97*(458), 632-648.

Innes, J. E., Booher, D. E., Di Vittorio, S. (2010). Strategies for Megaregion Governance. *Journal of the American Planning Association, 77,* 55-67. doi: 10.1080/01944363.2011.533640.Khadduri, J. (2005). Comment on Victoria Basolo and Mai Thi Nguyen's "Does mobility matter? The neighborhood conditions of housing voucher holders by race and ethnicity" *Housing Policy Debate, 16*, 325-334. doi: 10.1080/10511482.2005.9521547.

Kleinhans, R. and Varady, D. (2011). Moving Out and Going Down? A Review of Recent Evidence on Negative Spillover Effects of Housing Restructuring Programmes in the United States and the Netherlands, *International Journal of Housing Policy*, *11,* 155 – 174.

Le Gallo, J., & Chasco, C. (2012). Dealing with Data at Various Spatial Scales and Supports: An Application on Traffic Noise and Air Pollution Effects on Housing Prices with Multilevel

Models. In *Defining the Spatial Scale in Modern Regional Analysis* (pp. 281-309). Springer Berlin Heidelberg.

McClure, K. (2014). Housing Choice Voucher Location Patterns a Decade Later. *Housing Policy Debate,* doi: 10.1080/10511482.2014.921223.

McClure, K. (2008). Deconcentrating Poverty with Housing Programs. *Journal of the American Planning Association*, *74*, 90 - 99. doi: 10.1080/01944360701730165.

Mellow, J., Schlager, M. D., & Caplan, J. M. (2008). Using GIS to evaluate post-release prisoner services in Newark, New Jersey. *Journal of Criminal Justice*, *36*(5), 416-425.

Metraux, S., Caplan, J. M., Klugman, D., & Hadley, T. R. (2007). Assessing residential segregation among Medicaid recipients with psychiatric disability in Philadelphia. *Journal of Community Psychology*, *35*(2), 239-255.

Meltzer, R. (2013). Do Homeowners Associations Affect Citywide Segregation? Evidence From Florida Municipalities. *Housing Policy Debate*, *23*(4), 688-713.

Mínguez, R., Montero, J. M., & Fernández-Avilés, G. (2013). Measuring the impact of pollution on property prices in Madrid: objective versus subjective pollution indicators in spatial models. *Journal of Geographical Systems*, *15*(2), 169-191.

Mobley, L. R., Root, E., Anselin, L., Lozano-Gracia, N., & Koschinsky, J. (2006). Spatial analysis of elderly access to primary care services. *International Journal of Health Geographics*, *5*, 19. doi:10.1186/1476-072X-5-19.

Montero, J. M., Chasco, C., & Larraz, B. (2010). Building an environmental quality index for a big city: a spatial interpolation approach combined with a distance indicator. *Journal of Geographical Systems*, *12*(4), 435-459

Oakley, D. and Burchfield, K. (2009). Out of the Projects, Still in the Hood: The Spatial Constraints on Public-Housing Residents' Relocation in Chicago. *Journal of Urban Affairs, 31*, 589-614. doi: 10.1111/j.1467-9906.2009.00454.

O'Sullivan, D., & Unwin, D. (2014). Geographic information analysis, 3rd ed. John Wiley & Sons.

Páez, A. & Scott, D. (2004). Spatial statistics for urban analysis: a review of techniques with examples. *GeoJournal*, *61*, 53-67. doi: 10.1007/s10708-005-0877-5.

Páez, A., Farber, S., & Wheeler, D. (2011). A simulation-based study of geographically weighted regression as a method for investigating spatially varying relationships. *Environment and Planning-Part A*, *43*(12), 2992.

Patterson, K. and Yoo, E. E. (2012). Trapped in Poor Places? An Assessment of the Residential Spatial Patterns of Housing Choice Voucher Holders in 2004 and 2008. *Journal of Social Service Research, 38,* 637-655. doi: 10.1080/01488376.2012.706698.

Pendall, R. (2000). Why voucher and certificate users live in distressed neighborhoods. *Housing Policy Debate,* 11, 881-910. doi: 10.1080/10511482.2000.9521391.

Park, M. (2013). Housing vouchers as a means of poverty deconcentration and race desegregation: Patterns and factors of voucher recipients' spatial concentration in Cleveland. *Journal of Housing and the Built Environment*, 28, 451–468. doi 10.1007/s10901-012-9319-0.

Peeters, L., & Chasco, C. (2006). Ecological inference and spatial heterogeneity: an entropy-based distributionally weighted regression approach. *Papers in Regional Science*, *85*(2), 257-276.

Schafran, A. (2014). Rethinking Mega-Regions: Sub-Regional Politics in a Fragmented Metropolis. *Regional Studies, 48*, 587-602. doi**:** 10.1080/00343404.2013.834043.

Schafran, A. and Wegman, J. (2012). Restructuring, Race, and Real Estate: Changing Home Values and the New California Metropolis, 1989-2010, *Urban Geography*, *33,* 630-654. doi:10.2747/0272-3638.33.5.630.

Song, W. and Keeling, K. (2010). Location Patterns of Section 8 Housing in Jefferson County, Kentucky. *International Journal of Applied Geospatial Research, 1,* 1-18. doi: 10.4018/jagr.2010020901.

Soursourian, M. (2012). Suburbanization of Poverty in the San Francisco Bay Area. Federal Reserve Bank of San Francisco. Accessed on 10/18/2013. http://www.frbsf.org/community-development/files/Suburbanization-of-Poverty-in-the-Bay-Area1.pdf

Talen, E., & Koschinsky, J. (2014). The neighborhood quality of subsidized housing. *Journal of the American Planning Association, 80*(1), 67-82.

Tsutsumi, M., & Seya, H. (2008). Measuring the impact of large‐ scale transportation projects on land price using spatial statistical models. *Papers in regional science*, *87*(3), 385-401.

U.S. Census Bureau. ZIP Code Tabulation Areas (ZCTAs). Retrieved March 20, 2013, from http://www.census.gov/geo/reference/zctas.html.

U.S. Department of Housing and Urban Development. Housing Choice Vouchers Fact Sheet. Retrieved April 28, 2015, from http://portal.hud.gov/hudportal/HUD?src=/program_offices/public_indian_housing/programs/hcv/about/fact_sheet.

Varady, D. (2010). What Should Housing Vouchers Do? A Review of the Recent Literature. *Journal of Housing and the Built Environment*, *25,* 391-407. doi: 10.1007/s10901-010-9199-0.

Varady, D., Wang, X., Wang, Y., and Duhaney, P. (2010) The Geographic Concentration of Housing Vouchers, Blacks, and Poverty over Time: A Study of Cincinnati, Ohio, USA. *Urban Research and Practice, 3,* 39-62. doi: 10.1080/17535060903534172.

Wang, X., Varady, D., and Wang, Y. (2008) Measuring the Deconcentration of Housing Choice Voucher Program Recipients in Eight U.S. Metropolitan Areas Using Hot Spot Analysis. *Cityscape: A Journal of Policy Development and Research, 10,* 65-90.

Wang, X. and Varady, D. (2005) Using Hot-Spot Analysis to Study the Clustering of Section 8 Housing Voucher Families. *Housing Studies, 20,* 29-48. doi: 10.1080/0267303042000308714.

Wendland, N. (2015). All access: a micro-level case study on the secondary center of Berlin (1871–1936). *The Annals of Regional Science*, 1-25.

Winnick, L. (1995) The Triumph of Housing Allowance Programs: How a Fundamental Policy Conflict Was Resolved. *Cityscape: A Journal of Policy Development and Research,* 1, 99 – 121.

Wilson, R. (2012) Using Dual Kernel Density Estimation to Examine Changes in Voucher Density over Time. *Cityscape: A Journal of Policy Development and Research,* 14, 225-233.

Zhang, Q., Yang, Z., Li, Y., Chen, D., Zhang, J., and Chen, M. (2010) Spatial variability of soil nutrients and GIS-based nutrient management in Yongji County, China. *International Journal of Geographical Information Science*, 24, 965–981. doi: 10.1080/13658810903257954.

**Table 1. Growth rate and poverty by county (2010)**

| County | Growth rate | % living in poverty |
|---|---|---|
| Alameda | 4.61% | 11 |
| Contra Costa | 10.56% | 7.6 |
| Marin | 2.07% | 6.6 |
| Merced | 21.49% | 21.7 |
| Napa | 9.82% | 8.3 |
| San Francisco | 3.67% | 11.3 |
| San Mateo | 1.60% | 5.8 |
| Santa Clara | 5.89% | 7.5 |
| Solano | 4.77% | 8.3 |
| Sonoma | 5.51% | 8.1 |
| Sacramento | 15.96% | 14.1 |
| San Joaquin | 21.59% | 17.7 |
| Stanislaus | 15.09% | 16 |
| Yolo | 19.09% | 18.4 |

**Table 2: Spatial autocorrelation test results**[xxiv]

| | Queen Matrix | | 6 Nearest Neighbors Matrix | |
|---|---|---|---|---|
| | **Moran's I** | **Geary's c** | **Moran's I** | **Geary's c** |
| 2000 percentage of HCV households | 0.327 | 0.678 | 0.330 | 0.566 |
| 2010 percentage of HCV households | 0.301 | 0.706 | 0.292 | 0.603 |

**Table 3: Summary of the six ordinary kriging models for 'Percentage of HCV households'**

| | Model | RMSS | MS | ASE | RMS | RMS - ASE |
|---|---|---|---|---|---|---|
| **2000 (n = 6,155)** | Spherical | **0.9695** | 0.0028 | 0.0261 | 0.0249 | -0.0012 |
| | Rational Quadratic | 0.9742 | **0.0027** | **0.0259** | **0.0249** | **-0.0010** |
| | Gaussian | 0.9704 | 0.0029 | 0.0261 | 0.0249 | -0.0012 |
| **2010 (n = 6,636)** | Spherical | 0.8483 | 0.0014 | 0.0234 | 0.0197 | -0.0037 |
| | Rational Quadratic | **0.8276** | 0.0018 | 0.0239 | 0.0197 | -0.0042 |
| | Gaussian | 0.8474 | **0.0014** | **0.0234** | **0.197** | **-0.0037** |

***Note:*** *RMSS, Root Mean Square Standardized; MS, Mean Standardized; ASE, Average Standard Error; RMS, Root Mean Square; RMS – ASE, Difference between Root Mean Square and Average Standard Error*

**Table 4: Descriptive statistics of variables used in the regressions**

| Variable | Description | Mean | Minimum | Maximum | St. Dv. |
|---|---|---|---|---|---|
| **PERC_HCV** | Percentage of HCV households | 0.016 | 0 | 0.35 | 0.026 |
| **PER_AFR** | Percentage of total population reporting as African American or Black race | 8.33 | 0.00 | 100 | 12.94 |
| **PER_GRA** | Percentage of population 25 years and older with graduate degree | 12.05 | 0.00 | 75.60 | 11.39 |
| **PER_POV** | Percentage of total population with Income in 1999 below poverty level | 10.97 | 0.00 | 100 | 10.94 |
| **HOUSE** | Median value (dollars) for all owner-occupied housing units | 307,053 | 0.00 | 1,000,000 | 212,999 |

**Table 5: Pearson Coefficient for all variables used in the regressions**

| | % HCV households | % African American | Median housing price | % with graduate degrees | % in poverty |
|---|---|---|---|---|---|
| **% HCV households** | 1.00 | | | | |
| **% African American** | 0.323 | 1.00 | | | |
| **Median housing price** | -0.256 | -0.296 | 1.00 | | |
| **% with graduate degrees** | -0.261 | -0.245 | 0.709 | 1.00 | |
| **% in poverty** | 0.257 | 0.387 | -0.408 | -0.356 | 1.00 |

**Table 6: OLS and Spatial regressions results**

| Variables | OLS | Spatial Lag | Spatial Error |
|---|---|---|---|
| Constant | 0.0153*** | 0.008*** | 0.015*** |
| PER_POV | 0.0002*** | 0.0002*** | 0.0002*** |
| PER_AFR | 0.0005*** | 0.00003*** | 0.0005*** |
| PER_GRA | -0.0003*** | -0.0002** | -0.0002*** |
| HOUSE | -0.00000006*** | -0.00000004** | -0.00000001*** |
| Lag variable | ___ | 0.401*** | ___ |
| Lambda | ___ | ___ | 0.426*** |
| Adjusted R-squared | 0.149 | ___ | ___ |
| Log Likelihood | 14,138 | 14,419 | 14,444 |
| Akaike Criterion | -28,267 | -28,826 | -28,878 |
| Schwartz Criterion | -28,234 | -28,785 | -28,844 |
| Lagrange Multiplier (lag) | 761*** | ___ | ___ |
| Robust LM (lag) | 3.39*** | ___ | ___ |
| Lagrange Multiplier (error) | 835*** | ___ | ___ |
| Robust LM (error) | 77*** | ___ | ___ |

Note: *** p<0.01, ** p<0.05, * p<0.1, standard errors in parentheses

Figure 1: Core and periphery counties in the San Francisco Bay region

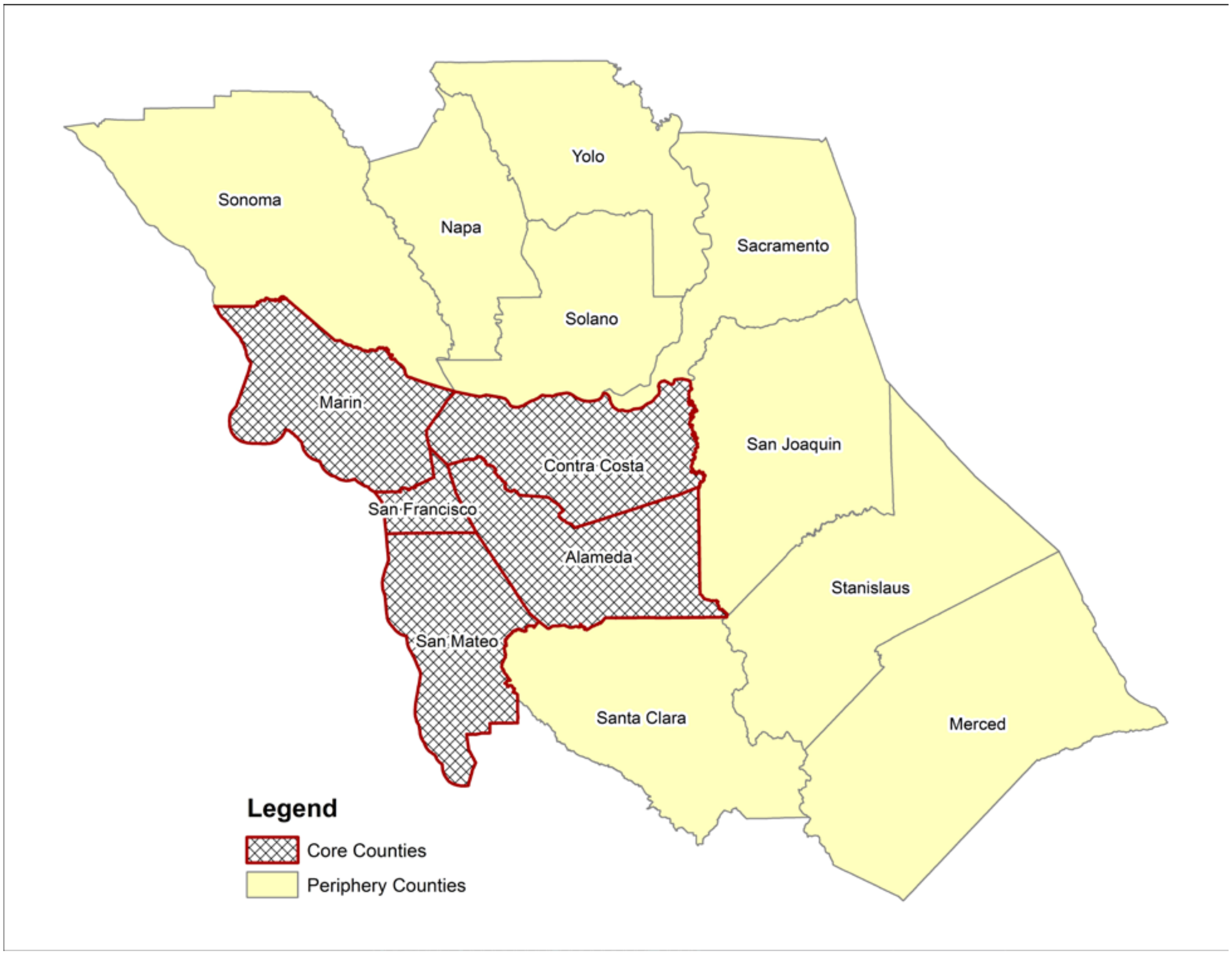


Figure 2: Aggregation effect in change of support system problem

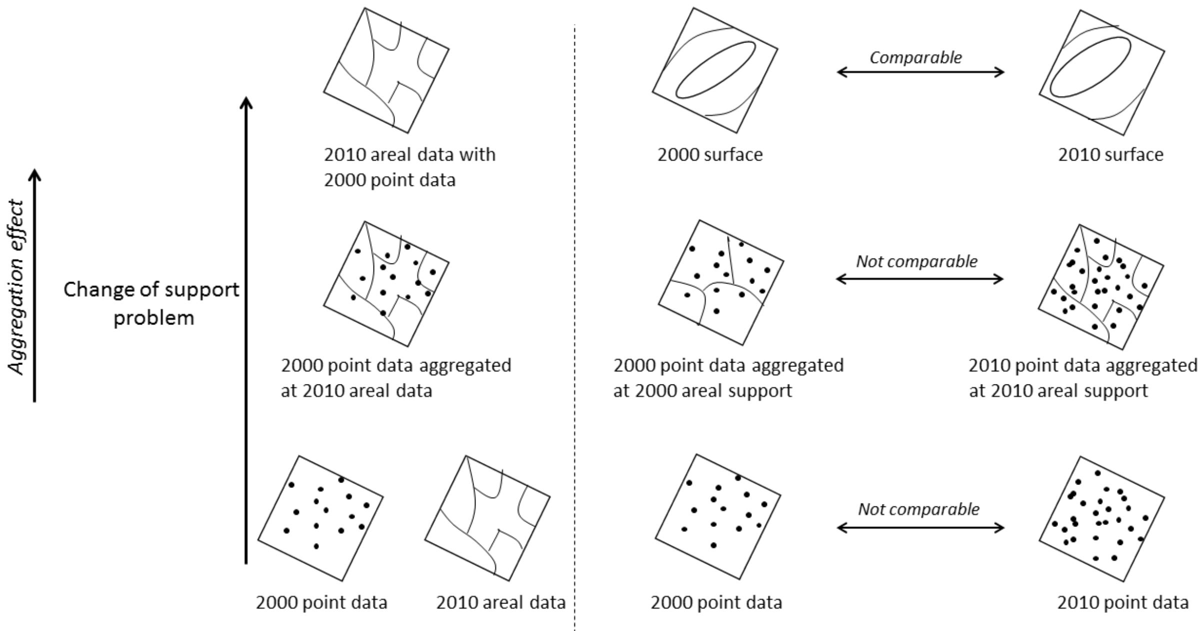



Figure 3: Ordinary kriging surfaces based on percentage of HCV households

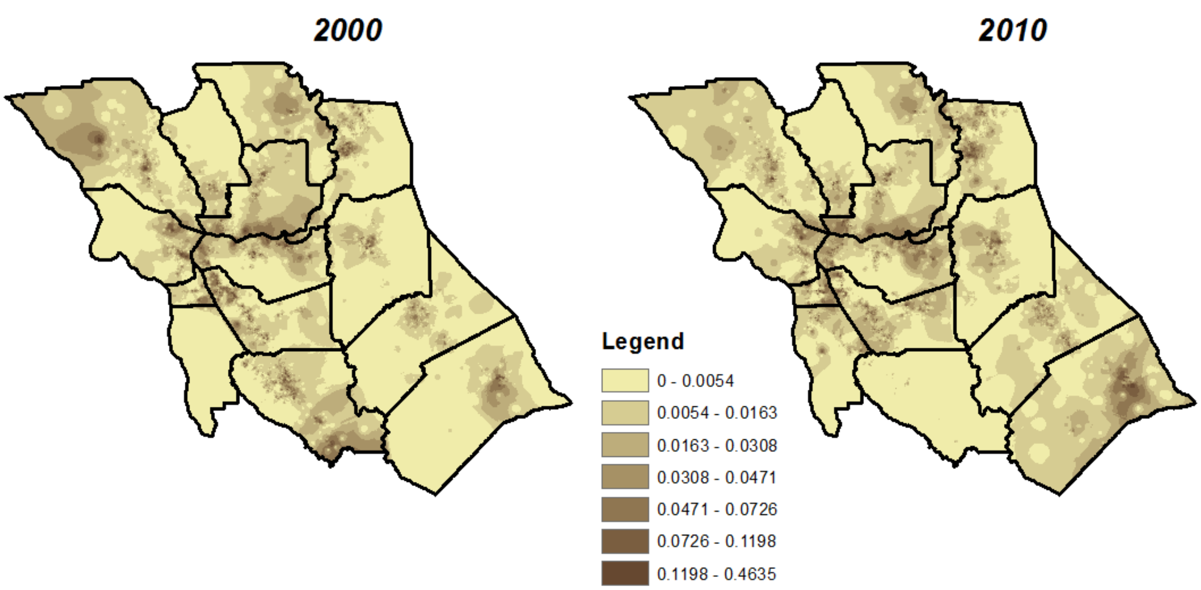


Figure 4: 2010 kriging surface minus 2000 kriging surface

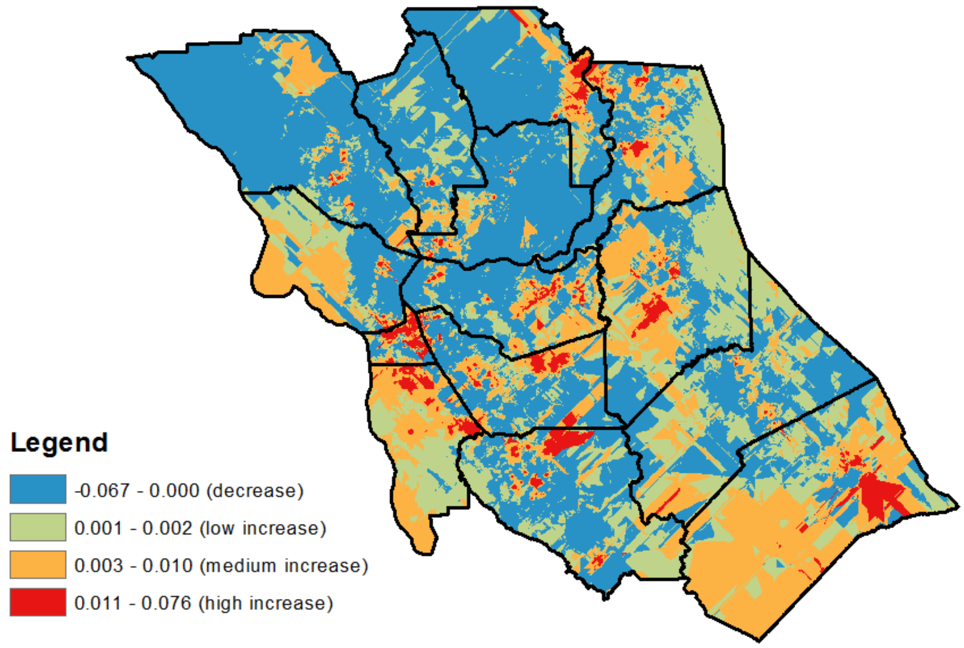


Figure 5: Kernel density surfaces for percentage of HCV households

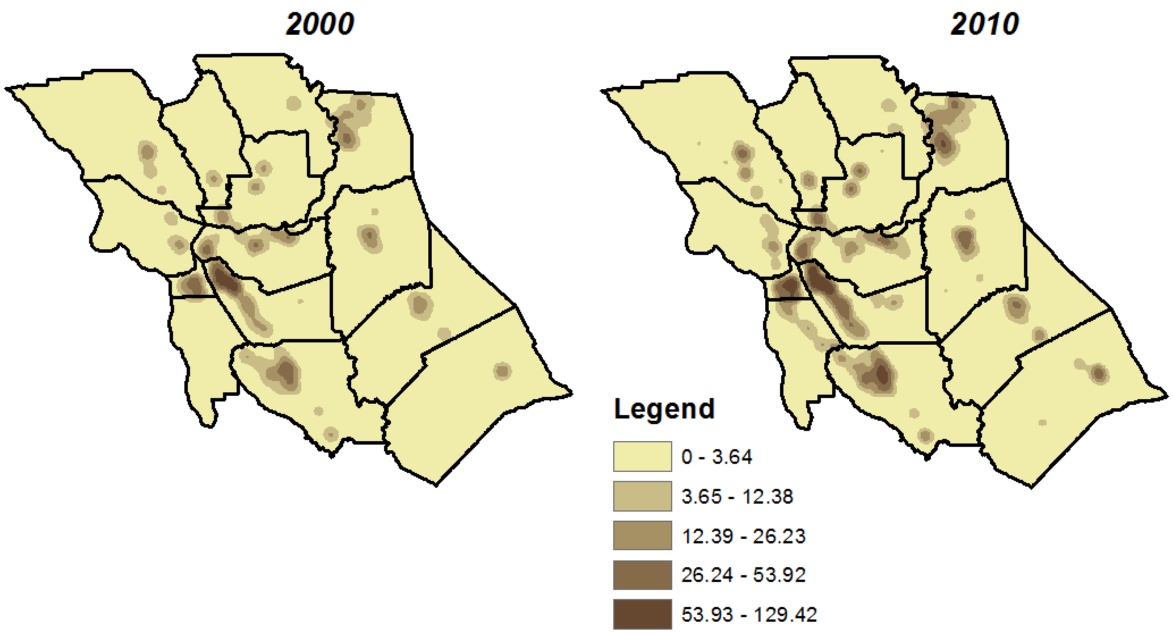
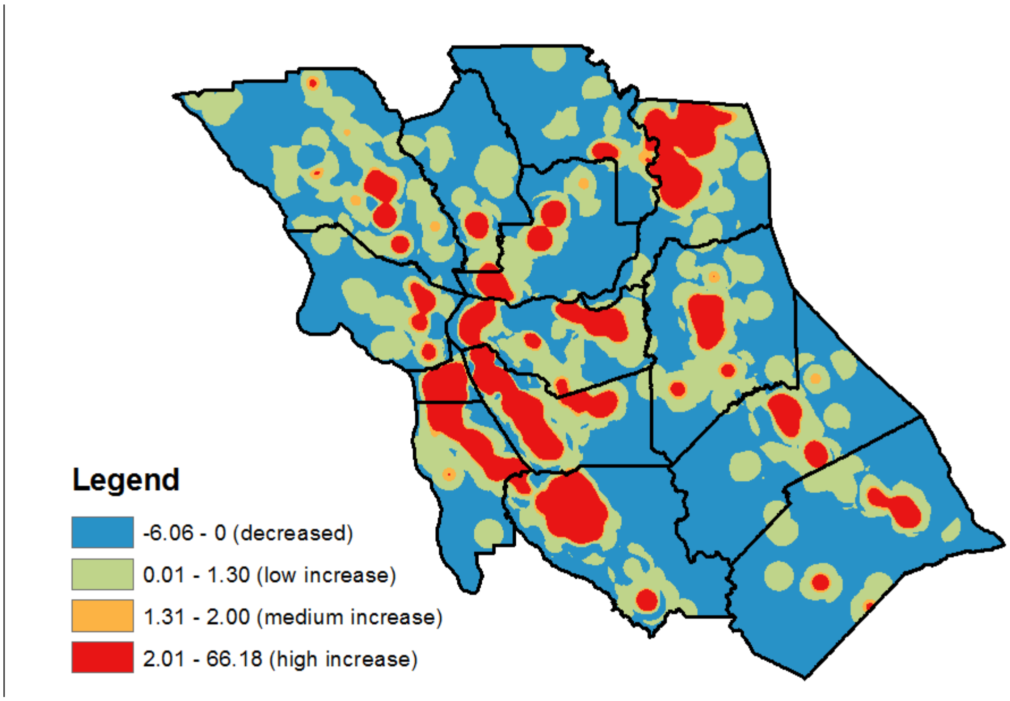


Figure 6: 2010 kernel density minus 2000 kernel density




## Notes

---

i "Millions of Americans Spend Over 30 Percent of Income for Housing" Downloaded on 12/2/2014. http://harvard-cga.maps.arcgis.com/apps/StorytellingTextLegend/index.html?appid=18d215ddb20946a4a16ae43586bf0b52

ii To measure poverty, the US Census Bureau uses "a set of money income thresholds that vary by family size and composition to determine who is in poverty. If a family's total income is less than the family's threshold, then that family and every individual in it is considered in poverty." That threshold is sometimes referred to as "the poverty line." The most appropriate method for measuring poverty has been a topic of significant debate. However, unless otherwise noted, this paper uses poverty data from the US Census and, therefore, the US Census Bureau's definition of poverty.

iii For more extensive reviews of the current research regarding HCV distributions and household mobility see Varady, 2010 or Galvez, 2010.

iv The Federal Reserve Bank of San Francisco's research brief on the suburbanization of poverty in the San Francisco Bay Area uses the Bay Area's Metropolitan Transportation Commission's definition of an urban census tract. Soursourian writes, "According to their guidelines, an 'urban' census tract has a population density greater than 10,000 persons per square mile (p. 13)." The Federal Reserve study "uses a binary delineation between urban and suburban, with 10,000 persons per square mile as the threshold (p. 13)."

[v] This database contains the HUD Family Report (HUD Form 50058) records reported by local public housing authorities (PHAs). The reports are filed at lease up and annually (with exceptions for MTW housing authorities) thereafter.

[vi] The variable was calculated by dividing the number of HCVs in each block group by the total number of HCV households in whole study area and then multiplied by 100.

[vii] Because kriging has been most commonly applied to biophysical topics, the use of this technique within the social sciences is not without controversy. Beeco and Brown state, "It can be difficult to adapt [kriging] techniques to the structure of social science data. (2013, p. 80)." Siniscalchi et al suggest that one reason for this difficulty is that social data is more likely than biophysical data to be characterized by high variability; e.g., one point can have a high value HCV and its neighbor point can have a low HCV value (2006). However, as a spatial interpolation technique, kriging is a powerful tool with established statistical rigor that enables the user to account for spatial dependency within a dataset, address change of support problems, and compare surfaces from different years. As a result, we have chosen to use this approach in our study.

[viii] The Moran's $I$ is a statistic that measures global spatial autocorrelation, providing a formal indication of the degree of linear association between the observed values and the spatially weighted averages of neighboring values. It is widely used in housing studies (Morckel, 2014; Zou, 2014; Saphores and Li, 2012). Geary's $c$ also measures global spatial autocorrelation, but it is not easily available in current spatial statistics software. We used SpaceStat to compute Geary's $c$.

[ix] In this study, k equals to six, representing the most frequent number of neighbors, based on the queen contiguity matrix. The simple binary queen contiguity matrix is composed of 0 and 1: if census block group $i$ has a common boundary and/or vertex with census block group $j$, then they are neighbors and $wi\ j = 1$; if census block group $i$ does not have a common boundary and/or vertex with census block group $j$, then they are not neighbors and $wi\ j = 0$. The diagonal elements are set to 0. Both matrices mentioned above must be row-normalized, meaning that when the spatial weight matrix $W$ is calculated, each element in row $i$ will be divided by the sum of row $i$'s elements.

[x] The kriging spatial interpolator used in this analysis "is considered as the optimal spatial interpolation for making best linear unbiased estimates of regionalized variables at unknown locations (Zhang et al., 2010, p. 968)."

[xi] There are three types of kriging: simple kriging, ordinary kriging, and universal kriging. After comparing various interpolation methods for air quality in hedonic models, Anselin and Le Gallo (2006) concluded that ordinary kriging is the most reliable, "providing the best results in terms of estimates (signs), model fit and interpretation" (p. 31). As a result, we chose to use ordinary kriging for this study.

[xii] Kriging assumes that the variable under examination is normally distributed. Because the 'Percentage of HCV households' variable was not normally distributed, we transformed it using a natural log.

[xiii] To define the lag size for the variogram models, we used the six-nearest-neighbors matrices to determine the average distance for nearest neighbor. This distance was used as the lag size. To determine the number of lags for the variogram models, we identified the largest distance among all points, and assured that the number of lags multiplied by the average distance for nearest neighbor would not exceed half of the largest distance. For 2000, the lag size used in the variogram models was 5,878 meters and the number of lags was six. For 2010, the lag size used in the variogram models was 5,762 meters and the number of lags was five. By using the values described above for lag size and number of lags we were able to assure that spatial autocorrelation was not masked and that the average bins were representative for all models (ESRI 2015b).

[xiv] The best models were identified based on the following criteria: 1) Root Mean Square Standardized should approach one; 2) Mean Standardized should approach zero; 3) Average Standard Error should be as low as possible; 4) Root Mean Square should be as low as possible; and 5) Root Mean Square and Average Standard Error should converge (ESRI 2013c).

[xv] To compare 2000 and 2010, we needed shapefiles from each year with comparable extents. However, zipcode boundaries shift from one year to the next. To create the comparable extents, we transformed all ZCTAs from the 2000 and 2010 shapefiles into points based on polygon centroids. We then eliminated all points from the 2000 data that fell outside the 2010 ZCTA polygons. These final point representations were used in the analyses described below.

[xvi] The 2000 symbology, i.e., cutoff points in the intervals of the legend, were based on natural breaks. To make both years comparable, the 2010 legend was manually defined using the same cutoff points as 2000. Our intention is to allow the reader to visually examine the difference between both spatial distributions. Natural breaks "optimally assign data to classes such that the variances within all classes are minimized, while the variances among classes are

maximized. In this manner, the data distribution is explicitly considered for determining class breaks" (BRFSS, 2015).

[xvii] We used ArcGIS 10.2, including the extension Geostatistical Analyst, OpenGeoDa and SpaceStat software to complete both the kriging analysis and the kernel density analysis that follows.

[xviii] For a more complete explanation of the use of kernel density estimation tests to evaluate changes in voucher density over time see Wilson, 2012.

[xix] "Kernel density estimation delineates the density of points [percentage of HCV] as a smoothly curved surface. Peaks of the surface are areas where [percentage of HCV] densely locates, and valleys show areas of low density of [percentage of HCV] (Dai et al, 2013, p.13)."

[xx] As with the earlier figure comparing 2000 and 2010 data, the 2000 symbology, i.e., cutoff points in the intervals of the legend, were based on natural breaks. To make both years comparable, the 2010 legend was manually defined using the same cutoff points as 2000.

[xxi] This problem may arise when there are unobservable independent variables in the model specification. As a consequence, the probability of having HCV households and poverty rates in an area may be correlated to the unobservable variables. If both HCV households and poverty rates are correlated to unobservable variables, then the estimated coefficients in the regression models would be biased. Not only poverty rates, but also other independent variables may present endogeneity problem as a result of correlation with unobservable variables.

[xxii] In a study examining HCV distributions in Cleveland, Park (2013) also used these methodological steps to incorporate spatial dependence and spatial heterogeneity in his regressions.

[xxiii] We use spatial matrices to capture the spatial structure of the study area. These tests may lead to specifying a spatial lag model or a spatial error model, indicating that an OLS is inefficient and/or biased (Anselin, 1988).

[xxiv] The *p*-value for these tests was 0.001.