# Measuring Neighborhood Quality with AHS and CSS Data: A Bayesian Approach

**Brent D. Mast**

U.S. Department of Housing and Urban Development

Abstract

While neighborhood quality is important for public policy, it is also difficult to quantify. This study measures neighborhood quality using data from two sources: the 2002 American Housing Survey (AHS), and HUD's Customer Satisfaction Survey (CSS) of Section 8 Housing Choice Voucher Program (HCVP) households. Survey responses are analyzed regarding neighborhood quality, home quality, and crime perceptions. Tract level Bayesian estimates are computed using AHS metro level data and CSS census tract data.

Compared to estimates solely based on CSS data, the Bayesian estimates have fewer outliers. Bayesian analysis also allows for estimation for tracts with lower sample sizes than would be practical using only CSS data. The Bayesian estimates tend to correlate more strongly with these auxiliary variables, and the differences are more apparent for tracts with larger differences between the CSS and Bayesian estimates.

*The contents of this article are the views of the author and do not necessarily reflect the views or policies of the U.S. Department of Housing and Urban Development or the U.S. government.*

# I) Introduction

Measuring neighborhood quality is important for many public policies. For instance, HUD's Housing Choice Voucher Program HCVP is intended to expand social and economic opportunities "outside areas of poverty or minority concentration" (HUD 2008: 2-1). In other words, the program is designed to promote access to decent and affordable housing in higher quality neighborhoods compared to neighborhoods of traditional Section 8 housing projects.

Yet neighborhood quality is inherently complex and difficult to measure. Data are available on a wide variety of neighborhood characteristics, such as poverty rates, income, crime rates, and school test scores. And while many policy makers and researchers rely on such indicators, they may have limited ability to measure the quality of neighborhood life as rated by residents (Buron and Patrabansh, 2008).

Survey data are available that measure residents' subjective perceptions of their neighborhoods. This study analyzes neighborhood quality perception data from two surveys: the American Housing Survey (AHS) and HUD's Customer Satisfaction Survey (CSS).

The AHS collects a large amount of information on housing conditions of American households. [1] The AHS is actually two surveys, metro and national, taking place in different years. I employ 2002 metro data for this study.

HUD's Customer Satisfaction Survey (CSS) was a three year survey of Section 8 Housing Choice Voucher Program (HCVP) households. [2] Conducted between 2000 and 2002, its main objective was to provide independent housing quality data to public housing agencies. About 460,000 responses were collected.

While the AHS and CSS contain many related questions, survey design differences make direct comparison of AHS and CSS data difficult.[3] However, for a subset of data items, estimates from both surveys correlate well. For instance, despite substantial differences in question wording, Mast (2009a) reports similar crime perception estimates based on CSS and 2001 AHS data.

Two of the most similar questions ask CSS and AHS respondents to rank the home and neighborhood on a scale of 1-10. Estimates for HCVP households from the CSS and 2001 AHS are very close (Mast 2009b). While many studies have examined differences in estimates from independent surveys, few researchers have attempted to combined information from independent surveys with Bayesian methods. This study attempts to extend this literature by using Bayesian methods to produce neighborhood quality indicators based on the both the AHS and CSS.

While there are sharp ideological differences between Bayesian and classical (or frequentist) statistics, in practice the most important difference concerns the use of prior information.[4] While classical methods tend to let the data "speak for themselves", Bayesian estimates always condition on prior information. For this study, I start with prior information from the AHS and update these estimates with CSS data.

I employ a particular Bayesian approach, referred to as a Bayesian Hierarchical Model, using metro level AHS data and CSS census tract data. Compared to tract estimates solely based on CSS data, the Bayesian estimates have fewer outliers. By drawing strength from the AHS, Bayesian analysis also allows for inference with smaller samples than would be practical using only CSS data.

To validate my estimates, I examine correlation of the CSS and Bayesian estimates with other measures of neighborhood quality, such as median income, poverty rates, and indicators for tracts receiving Low Income Housing Tax Credits. The Bayesian estimates tend to correlate more strongly with these auxiliary variables, and the differences are more apparent for tracts with larger differences between the CSS and Bayesian estimates.

The remainder of the article proceeds as follows. The next section reviews relevant studies. The survey data are then described. The model is explained next, followed by empirical results. Estimates are then compared with other tract-level measures of neighborhood quality. The final section summarizes my results.

## II) Literature Review

*Measuring Neighborhood Quality* Neighborhood quality is a difficult concept to quantity. According to Dubin (1992), measurement error is a likely cause for the lack of consistent effects of neighborhood quality indicators in hedonic housing price regressions.

The stalwarts of neighborhood quality measurement have traditionally been data on income, race, ethnicity, and poverty. Yet until recently, reliable neighborhood level population, income, and poverty data were only available from the decennial Census.[5]

Crime rates may also be useful measures of neighborhood quality. For instance, Deller and Ottem (2001) use county crime rates as neighborhood quality controls in hedonic property value regressions.

Crime rate data are also available at lower levels of aggregation for some localities. Cahill (2006) reports crime rates (averaged over 1998-2002) for census tracts and block groups for three U.S. cities (Nashville, Tucson, and Portland).[6] An increasing number of areas, such as Seattle, are making neighborhood crime data available through there crime mapping programs.[7]

Neighborhood quality should also be positively related to educational achievement. Sedgley et al. (2008) find that 8$^{th}$ grade test scores and SAT scores have significant positive effects on housing prices. They find no consistent effect for 3$^{rd}$ grade scores, however.

Survey measures are available that measure residents' subjective perceptions of their neighborhoods. Buron and Patrabansh (2008) provide evidence that subjective perceptions of neighborhood quality may not correlate highly with objective measures, such as poverty rates.

A related literature studies differences in perceived neighborhood quality in the same localities due to differences in characteristics such as race, ethnicity, gender, and income (St. John and Clark 1984). Differences may be especially apparent regarding neighborhood crime (Austin et al. 2002). For instance, females may be more vulnerable to certain crimes, particularly rape. In addition, racial or ethnic minorities may have different attitudes towards crime and law enforcement due to racial or ethnic differences in arrest rates and incarceration rates.

Many researchers have measured perceived neighborhood quality with AHS data (Newman and Schnare 1993, Dilulio 1994, Chapman and Lombard 2006). For instance, Hipp (2007) studies the relationship between AHS household crime perceptions and county crime rates. He finds household perceptions of crime are more strongly related to violent crime than property crime.

Other studies have measured neighborhood quality with CSS data. Buron and Patrabansh (2008) study the relationship between CSS household neighborhood quality responses and census data. As indicated above, they found little correspondence. This calls into question use of social indicators such as poverty rates as measures of neighborhood quality.

Buron and Patrabansh's findings may be affected by their use household level data. This study finds that resident perceptions aggregated to the tract level have fairly strong correlation with poverty and income. They also limit their analysis to census variables. Their model might have more explanatory power if other variables were considered, such as crime rates.

Gray, Haley, and Mast (2009) report wide variation in CSS neighborhood ratings across demographic groups. Mast (2009b), using CSS data, estimates that West Virginia crime perceptions relate more strongly with property crime than violent crime.

*Comparing Estimates from Independent Surveys* Numerous studies have compared and contrasted estimates from independent surveys. For example, Bishaw and Stern (2006) examine differences in poverty estimates based on the American Community Survey (ACS) and Current Population Survey (CPS).

A few studies have compared CSS and AHS estimates. Buron, Kaul, and Patterson (2003) match 2001 CSS households with a sample of unassisted AHS households. While they report lower housing quality for HCVP households relative to similar unassisted families, they caution that their results may be driven by methodology and questionnaire differences.

Mast(2009a) studies crime perception questions on the CSS and 2001 AHS. The wording of the crime question differs on the two surveys. The AHS asks if there is "a neighborhood crime problem", while the CSS asks if crime or drugs "is a big problem in (the) neighborhood". Response options also differ. Despite these discrepancies, Mast(2009a) recodes responses into binary indicators with similar means. 31.5 % of AHS HCVP households are estimated to have a crime problem, compared to 33.8 % of CSS households.

Two of the most similar AHS and CSS questions ask respondents to rank their home and neighborhood on a scale of 1-10. Mast (2009b) compares both rankings on the CSS to those for HCVP households in the 2001 AHS. For both homes and neighborhoods, CSS rankings are just slightly higher than AHS estimates for HCVP homes.

According to Mast (2009a), because AHS and CSS crime estimates are similar, they are well suited for Bayesian methods. Since the CSS sample size is much larger than the AHS, he employs a Bayesian Hierarchical Model. AHS national estimates are used as priors along with CSS county data to estimate Bayesian posterior county estimates. Compared to estimates solely based on CSS data, the Bayesian estimates have lower variance and correlate more highly with county violent and property crime rates. Consistent with Hipp (2007), the relationship is strongest with violent crime.

## III) Data Description

In this section, I report 2002 AHS and CSS summary statistics on three measures of neighborhood quality: neighborhood ratings, home ratings, and crime perceptions.

Both surveys ask respondents to rate the neighborhoods and homes on an ordinal scale of 1-10. Because the response categories are numerical, we could compute mean ratings. However, we would be making an assumption that a rating of 6 is twice as great as a rating of 3. For ordinal data it is customary to only compute order statistics, such as the median or other percentiles.

Exhibit 1 reports percentiles (10th, 25th, median, 75th, and 90th) for neighborhood and home ratings. AHS data are reported for all occupied rental units and HCVP household. The CSS ratings are based on responses for HCVP households in the 13 AHS metro areas between 2000 and 2002.[8] Survey responses for AHS renters are weighted to be representative of all renters in the 13 metro areas; AHS HCVP and CSS responses are weighted to be representative of all voucher households in the 13 metro areas.[9]

Exhibit 1: Neighborhood and Home Rating Percentiles

| | Neighborhoods | | | Homes | | |
|---|---|---|---|---|---|---|
| Weighted Percentile | AHS - All Renters | AHS-HCVP Households | CSS | AHS - All Renters | AHS-HCVP Households | CSS |
| 10th Percentile | 5 | 5 | 4 | 5 | 5 | 4 |
| 25th Percentile | 6 | 6 | 6 | 6 | 7 | 6 |
| Median | 8 | 8 | 8 | 8 | 8 | 8 |
| 75th Percentile | 9 | 9 | 9 | 9 | 10 | 10 |
| 90th Percentile | 10 | 10 | 10 | 10 | 10 | 10 |

Source: author's calculation using 2002 AHS and CSS data. For neighborhood ratings N equals 16,458 for AHS renters, 503 for AHS HCVP households, and 26,822 for the CSS. For home ratings N equals 16,510 for AHS renters, 503 for AHS HCVP households, and 26,987 for the CSS.

Neighborhood ratings correspond highly with home ratings. 25th percentile neighborhood ratings are 6 for all three samples (AHS renters, AHS HCVP, and CSS). 25th percentile home ratings are 6 for AHS renters, 7 for AHS HCVP households, and 6 for CSS families. All median ratings are 8. 75th percentile neighborhood ratings are 9 for both AHS samples, and 10 for the CSS. 75th percentile home ratings are 9 for both AHS samples, and 10 for the CSS.

We can compute binary indicators of high neighborhood and home ratings for which mean analysis is appropriate. To demonstrate, for this study I will treat ratings of at least 8 as high ratings. The downside of this approach is that the threshold for high ratings is somewhat arbitrary.

Exhibit 2 reports mean percentages of households with high neighborhood and high home ratings. More than half of households in each sample rate their neighborhood 8 or above. 55.6% of AHS renters have high neighborhood ratings, as do 54.4% of AHS HCVP households, and 52.8% of CSS households.

Exhibit 2: Mean Indicators of High Neighborhood Quality

| | AHS - All Renters | | AHS - HCVP Households | | CSS - HCVP Households | |
|---|---|---|---|---|---|---|
| Variable | Responses | Weighted Mean | Responses | Weighted Mean | Responses | Weighted Mean |
| High neighborhood rating | 16458 | 0.556 | 503 | 0.544 | 26822 | 0.528 |
| High home rating | 16510 | 0.547 | 503 | 0.644 | 26987 | 0.596 |
| Low crime indicator | 16777 | 0.770 | 509 | 0.677 | 27376 | 0.664 |

High home and neighborhood ratings are ≥8 on a 1-10 scale. Source: author's calculation using 2002 AHS and CSS data.

On average, voucher families tend to rate their homes better than renters in general. 64.4% of AHS HCVP families rate their homes 8 or above, as do 59.6% of CSS households; the corresponding mean for all AHS renters is 54.7%.

The wording of the crime question differs on the two surveys. The AHS asks households if their "neighborhood has a neighborhood crime problem". Response categories include "No", "Don't Know", and "Yes". The CSS asks if crime or drugs "is a big problem in (the) neighborhood." Response categories include "No Problem", "Don't Know", "Some Problem", and "Big Problem".

To facilitate comparison of crime variables from both surveys, I recode responses as binary indicators of low crime. For the AHS, "Yes" responses are set to zero, while "No" and "Don't know" responses are treated as ones. For the CSS, "some problem" and "big problem" responses are set to zero, and "no problem" and "don't know" responses are set to one. Non-responses for both surveys are set to missing.

Mean indicators of low crime perceptions are reported in Exhibit 2. Compared to all renters, voucher households rate the neighborhoods as less safe. 77.0% of all renters do not perceive a major crime problem in their area. The corresponding means are 67.7% for AHS voucher households, and 66.5% of CSS households.

Exhibits 3-5 report summary statistics by metro area for high neighborhood ratings, high home ratings, and low crime indicators, respectively. According to the AHS, Phoenix has the lowest fraction of high neighborhood rating for all renters (.517) and HCVP households (.322). The Kansas City metro area has the best neighborhood ratings according to all AHS renters (.595), and the worst neighborhoods according to CSS households. One possible explanation for differing opinions between all renters and voucher households is that affordable rental units meeting HUD housing quality guidelines in the Kansas City area are more prevalent in lower quality neighborhoods. According to voucher households participating in the CSS, Santa Ana has the best neighborhoods.

Exhibit 3: Mean Indicators of High Neighborhood Ratings, by Metro Area

| | AHS - All Renters | | AHS - HCVP Households | | CSS - HCVP Households | |
|---|---|---|---|---|---|---|
| Metro Area | Responses | Weighted Mean | Responses | Weighted Mean | Responses | Weighted Mean |
| Santa Ana-Anaheim-Irvine, CA Division | 1513 | 0.562 | 44 | 0.563 | 1624 | 0.610 |
| Buffalo-Cheektowaga-Tonawanda, NY | 1022 | 0.535 | 57 | 0.485 | 1734 | 0.546 |
| Charlotte-Gastonia-Concord, NC-SC | 1014 | 0.579 | 18 | 0.600 | 2237 | 0.538 |
| Columbus, OH | 1230 | 0.541 | 55 | 0.464 | 1651 | 0.453 |
| Dallas-Plano-Irving, TX Division | 1437 | 0.560 | 41 | 0.616 | 3677 | 0.494 |
| Fort Worth-Arlington, TX Division | 1203 | 0.526 | 40 | 0.523 | 1584 | 0.498 |
| Kansas City, MO-KS | 1058 | 0.595 | 43 | 0.488 | 2938 | 0.442 |
| Miami-Fort Lauderdale-Miami Beach, FL | 1218 | 0.574 | 27 | 0.590 | 2807 | 0.591 |
| Milwaukee-Waukesha-West Allis, WI | 1388 | 0.581 | 30 | 0.566 | 1681 | 0.451 |
| Phoenix-Mesa-Scottsdale, AZ | 1200 | 0.517 | 30 | 0.322 | 2375 | 0.499 |
| Portland-Vancouver-Beaverton, OR-WA | 1271 | 0.525 | 32 | 0.433 | 1609 | 0.534 |
| Riverside-San Bernardino-Ontario, CA | 1386 | 0.553 | 28 | 0.563 | 1344 | 0.514 |
| San Diego-Carlsbad-San Marcos, CA | 1518 | 0.570 | 58 | 0.677 | 1561 | 0.568 |

High neighborhood ratings are ≥8 on a 1-10 scale. Source: author's calculations using CSS and 2002 AHS data.

Exhibit 4: Mean Indicators of High Home Ratings, by Metro Area

| | AHS - All Renters | | AHS - HCVP Households | | CSS - HCVP Households | |
|---|---|---|---|---|---|---|
| Metro Area | Responses | Weighted Mean | Responses | Weighted Mean | Responses | Weighted Mean |
| Santa Ana-Anaheim-Irvine, CA Division | 1517 | 0.531 | 44 | 0.666 | 1641 | 0.701 |
| Buffalo-Cheektowaga-Tonawanda, NY | 1026 | 0.563 | 57 | 0.674 | 1741 | 0.608 |
| Charlotte-Gastonia-Concord, NC-SC | 1016 | 0.550 | 18 | 0.718 | 2242 | 0.583 |
| Columbus, OH | 1235 | 0.511 | 55 | 0.576 | 1660 | 0.526 |
| Dallas-Plano-Irving, TX Division | 1445 | 0.558 | 41 | 0.741 | 3690 | 0.539 |
| Fort Worth-Arlington, TX Division | 1203 | 0.509 | 40 | 0.640 | 1593 | 0.564 |
| Kansas City, MO-KS | 1060 | 0.589 | 43 | 0.580 | 2956 | 0.522 |
| Miami-Fort Lauderdale-Miami Beach, FL | 1218 | 0.546 | 27 | 0.670 | 2830 | 0.659 |
| Milwaukee-Waukesha-West Allis, WI | 1393 | 0.589 | 30 | 0.678 | 1691 | 0.543 |
| Phoenix-Mesa-Scottsdale, AZ | 1205 | 0.549 | 30 | 0.702 | 2390 | 0.547 |
| Portland-Vancouver-Beaverton, OR-WA | 1279 | 0.512 | 32 | 0.648 | 1630 | 0.624 |
| Riverside-San Bernardino-Ontario, CA | 1392 | 0.554 | 28 | 0.423 | 1356 | 0.587 |
| San Diego-Carlsbad-San Marcos, CA | 1521 | 0.549 | 58 | 0.639 | 1567 | 0.634 |

High home ratings are ≥8 on a 1-10 scale. Source: author's calculations using CSS and 2002 AHS data.

Exhibit 5: Mean Indicators of Low Crime Perceptions, by Metro Area

| | AHS - All Renters | | AHS - HCVP Households | | CSS - HCVP Households | |
|---|---|---|---|---|---|---|
| Metro Area | Responses | Weighted Mean | Responses | Weighted Mean | Responses | Weighted Mean |
| Santa Ana-Anaheim-Irvine, CA Division | 1530 | 0.843 | 44 | 0.865 | 1668 | 0.806 |
| Buffalo-Cheektowaga-Tonawanda, NY | 1074 | 0.783 | 60 | 0.745 | 1756 | 0.636 |
| Charlotte-Gastonia-Concord, NC-SC | 1028 | 0.806 | 19 | 0.659 | 2264 | 0.634 |
| Columbus, OH | 1243 | 0.686 | 56 | 0.530 | 1692 | 0.540 |
| Dallas-Plano-Irving, TX Division | 1461 | 0.725 | 42 | 0.668 | 3746 | 0.629 |
| Fort Worth-Arlington, TX Division | 1220 | 0.788 | 40 | 0.727 | 1627 | 0.638 |
| Kansas City, MO-KS | 1089 | 0.803 | 43 | 0.649 | 2981 | 0.598 |
| Miami-Fort Lauderdale-Miami Beach, FL | 1238 | 0.860 | 27 | 0.847 | 2876 | 0.729 |
| Milwaukee-Waukesha-West Allis, WI | 1418 | 0.725 | 30 | 0.600 | 1713 | 0.591 |
| Phoenix-Mesa-Scottsdale, AZ | 1215 | 0.706 | 30 | 0.518 | 2418 | 0.572 |
| Portland-Vancouver-Beaverton, OR-WA | 1300 | 0.707 | 32 | 0.561 | 1642 | 0.703 |
| Riverside-San Bernardino-Ontario, CA | 1399 | 0.771 | 28 | 0.679 | 1379 | 0.684 |
| San Diego-Carlsbad-San Marcos, CA | 1562 | 0.763 | 58 | 0.645 | 1614 | 0.689 |

Source: author's calculations using CSS and 2002 AHS data.

Two of the three survey estimates rank Santa Ana as the safest metro area. 84.3% of Santa Ana AHS renters report no major crime problem, as do 86.5% of AHS HCVP households and 80.6 % of CSS respondents. Columbus is the least safe metro area according to two of three survey measures. 68.6% of AHS renters report no major crime problems in the Columbus area, compared to 53.0% of AHS HCVP households, and 54.0% of CSS households. According to AHS voucher respondents, the Phoenix area has the greatest perceived crime problem.

Exhibits 6 report summarizes absolute percentage differences between the metro level AHS and CSS means reported in exhibits 3-5. For each indicator of neighborhood quality, two differences are summarized– one between AHS renter means and CSS means, and another between AHS HCVP means and CSS means.

Differences in mean neighborhood indicators based on AHS renters range from .393% in Phoenix to 29.515% in Kansas City, with a median of 7.216% in Riverside. Neighborhood indicator differences based on the AHS HCVP sample are slightly larger on average. The mean difference based on AHS renters is 9.497%, versus 14.064% for the AHS HCVP sample. Differences based on the AHS voucher sample vary from .186% in Miami to 43.146% in Phoenix, with a median difference of 10.890% in Charlotte.

Differences in home indicators by metro area are quite similar in magnitude to differences in neighborhood indicators. The mean difference in home indicator means is 10.523% based on AHS renters, and 14.287% based on AHS vouchers.

Exhibit 6: Absolute Percentage Differences between AHS and CSS Metro Means

| | High neighborhood rating indicator | | High home rating indicator | | Low crime indicator | |
|---|---|---|---|---|---|---|
| Absolute % difference | AHS renters vs CSS | AHS | CSS | vs CSS | AHS vs CSS | AHS HCVP vs CSS |
| Minimum difference | 0.393% | 0.186% | 0.482% | 0.804% | 0.528% | 0.667% |
| Median difference | 7.216% | 10.890% | 8.010% | 10.457% | 20.449% | 7.025% |
| Mean difference | 9.497% | 14.064% | 10.523% | 14.287% | 16.758% | 8.622% |
| Maximum difference | 29.515% | 43.146% | 27.590% | 32.500% | 29.237% | 22.588% |

Source: author's calculations using CSS and 2002 AHS data.

Regarding crime, there is more agreement between the CSS and the AHS HCVP sample. Based on all AHS renters, differences in mean low crime indicators vary from .528% in Portland to 29.237% in Kansas City; the median difference is 20.449% in Milwaukee. Corresponding differences based on AHS HCVP households range from .667% for Riverside to 22.558% for Portland, with a median of 7.025% in the Santa Ana area. The Mean difference based on all AHS renters is 16.758%, compared to 8.622% for AHS HCVP households.

While exploring differences between the AHS and CSS estimates might make for an interesting study, this is not my focus. When independent surveys estimate the same variable, Bayesian methods can produce a more reliable estimate using information from both surveys. In the next section, I demonstrate how a Bayesian Hierarchical Model can produce more robust tract level estimates using metro AHS data and tract CSS data.

# IV) Data Analysis

In this section I analyze neighborhood, home, and crime measures from the CSS and 2002 AHS. My goal is to produce Bayesian tract estimates of neighborhood quality based on both surveys. While Conventional Bayesian updating would require AHS and CSS tract level estimates, the AHS sample is not large enough to produce reliable tract estimates. Therefore I choose a Bayesian Hierarchical Model using AHS metro estimates and CSS tract estimates.

AHS responses are aggregated at the metro level for 13 metro areas. 26,264 CSS responses are aggregated into 3,749 census tracts in the AHS metro areas. 2,397 CSS responses are excluded where either 1) the address not could not be accurately geocoded at the tract level, or 2) there was not a valid response for the home rating, neighborhood rating, or crime question.

*Neighborhood Indicators* While the neighborhood and home ratings are ordinal, such data do not easily lend themselves to Bayesian methods.[10] For my analysis, I use the binary indicators of high neighborhood ratings ($X_1$), high home ratings ($X_2$), and low crime perceptions ($X_3$) discussed above in Section III.

I also employ three composite variables. $X_4$ is a binary indicator for households with a high neighborhood rating and a high home rating. $X_5$ indicates a high neighborhood rating and a low crime perception. $X_6$ indicates a high neighborhood rating, a high home rating, and a low crime perception.

Only a small percent of the CSS census tract samples meet the usual normality criteria for any of the six indicators.[11] Therefore I assume $X_1$-$X_6$ follow a Binomial(n,$p_i$) distribution, for i=1 to 6. n represents the number of weighted responses, which is the same for all indicators in a given census tract. $p_i$ represents the probability that indicator $X_i$ equals 1. While each indicator has a separate distribution for each tract, for simplicity I do not use tract subscripts.

$p_i$ follows a Beta($\alpha_i$,$\beta_i$) probability distribution, where $\alpha_i$ equals the weighted count of high quality indicators. $\beta_i$ equals the weighted count of low quality indicators, which equals n- $\alpha_i$. The Beta probability distribution has mean $\alpha/(\alpha + \beta)$ and standard deviation equal to the square root of $\alpha\beta/[(\alpha + \beta)^2(\alpha + \beta +1)]$.

Using weighted counts based on original sampling weights summing to total HCVP households would treat estimated counts as known values. This would grossly understate variance by ignoring sampling variability. To reduce bias, I use weighted counts with adjusted weights summing to responses.[12] Compared to estimates based on the original sampling weights, this reweighting produces estimates with the same weighted means and more realistic variance.

Exhibit 7 reports descriptive statistics for the 3,749 CSS tract distributions of $X_1$-$X_6$. The 1st variable listed in Exhibit 7 is $X_1$, the indicator for high neighborhood ratings. Weighted responses for $X_1$ vary from .096 to 173.486, with a median of 3.984 and mean of 7.043. $\alpha_1$, the count of high neighborhood ratings, ranges from 0 to 171.083. The median number of high neighborhood ratings equals 2.088, and the mean is 3.720. $p_1$, the mean probability of a high neighborhood rating, varies across census tracts from 0 to 1. 543 tracts have $p_1$ =0 (no high neighborhood ratings), and 1,122 have $p_1$ =1 (all high ratings); these tracts have 0 standard deviation. The median probability of a high neighborhood rating is .619, and mean is .593.

CSS respondents rate their homes slightly higher on average than their neighborhoods. $X_2$ is the indicator of high home ratings, and the $p_2$ is the mean probability of a high home rating. The median value of $p_2$ is .680, and the mean is .638.

The majority of CSS households do not report major crime problems in their neighborhoods. $X_3$ indicates low crime perceptions, and $p_3$ is the mean probability of a low crime perception. $p_3$ has a median of .816, and a mean of .713.

$X_4$ indicates both a high neighborhood rating and a high home rating; $p_3$ indicates the mean probability that $X_4$=1. $X_5$ indicates a high neighborhood rating and low crime perception, and $p_5$ indicates the mean probability that $X_5$=1. $p_4$ has a mean of .527, and $p_5$ has a mean of .524. Rounded to three decimal points, $p_4$ and $p_5$ have the same median (.500).

$X_6$ indicates a high neighborhood rating, a high home rating, and a low crime perception. $p_6$ indicates the mean probability that $X_6$=1. Because $X_6$ is the most restrictive of the 6 indicators, it has the lowest median (.448) and mean (.471).

Exhibit 7: CSS Census Tract Summary Statistics

| | X1 (high neighborhood rating) | | | | X2 (high home rating) | | | |
|---|---|---|---|---|---|---|---|---|
| | Minimum | Median | Mean | Maximum | Minimum | Median | Mean | Maximum |
| Weighted Responses | 0.096 | 3.984 | 7.043 | 173.486 | 0.096 | 3.984 | 7.043 | 173.486 |
| Count of high ratings (α) | 0.000 | 2.088 | 3.720 | 171.083 | 0.000 | 2.281 | 4.202 | 171.901 |
| Count of low ratings (β) | 0.000 | 1.305 | 3.323 | 170.264 | 0.000 | 1.133 | 2.841 | 46.045 |
| Mean probability of a high rating (p) | 0.000 | 0.619 | 0.593 | 1.000 | 0.000 | 0.680 | 0.638 | 1.000 |
| Standard deviation of p | 0.000 | 0.087 | 0.093 | 0.443 | 0.000 | 0.091 | 0.094 | 0.443 |
| | X3 (low crime perception) | | | | X4 (high neighborhood and home ratings) | | | |
| | Minimum | Median | Mean | Maximum | Minimum | Median | Mean | Maximum |
| Weighted Responses | 0.096 | 3.984 | 7.043 | 173.486 | 0.096 | 3.984 | 7.043 | 173.486 |
| Count of high ratings (α) | 0.000 | 2.550 | 4.657 | 169.881 | 0.000 | 1.830 | 3.308 | 171.083 |
| Count of low ratings (β) | 0.000 | 0.662 | 2.386 | 45.114 | 0.000 | 1.587 | 3.735 | 170.264 |
| Mean probability of a high rating (p) | 0.000 | 0.816 | 0.713 | 1.000 | 0.000 | 0.500 | 0.527 | 1.000 |
| Standard deviation of p | 0.000 | 0.056 | 0.081 | 0.457 | 0.000 | 0.090 | 0.096 | 0.443 |
| | X5 (high neighborhood rating, low crime perception) | | | | X6 (high neighborhood and home ratings, low crime perception) | | | |
| | Minimum | Median | Mean | Maximum | Minimum | Median | Mean | Maximum |
| Weighted Responses | 0.096 | 3.984 | 7.043 | 173.486 | 0.096 | 3.984 | 7.043 | 173.486 |
| Count of high ratings (α) | 0.000 | 1.820 | 3.215 | 169.063 | 0.000 | 1.553 | 2.905 | 169.063 |
| Count of low ratings (β) | 0.000 | 1.581 | 3.828 | 171.466 | 0.000 | 1.847 | 4.138 | 171.466 |
| Mean probability of a high rating (p) | 0.000 | 0.500 | 0.524 | 1.000 | 0.000 | 0.448 | 0.471 | 1.000 |
| Standard deviation of p | 0.000 | 0.085 | 0.094 | 0.457 | 0.000 | 0.083 | 0.094 | 0.457 |

N=3,749 tracts. Source: author's calculations using CSS data.

*Bayesian Estimates* The Bayesian posterior distribution of $p_i$ for each tract follows a Beta($\alpha_i$*,$\beta_i$*) distribution where $\alpha_i$*= $\alpha_{i,prior}$ + $\alpha_i$, and $\beta_i$*= $\beta i,_{prior}$ + $\beta_i$. $\alpha_{i,prior}$ is our prior best guess for the number of high ratings, with no knowledge of the CSS data. $\beta_{i,prior}$ is our prior guess for the number low ratings. $\alpha_i$ is the CSS weighted count of high ratings, and $\beta_i$ is the CSS weighted count of low ratings.

Reliable tract level information from the AHS is not available. Therefore I employ a Bayesian Hierarchical Model adopted from Gelman et al. (2004: 131-135).[13] For each metro area, I use a common prior distribution for all tracts based on the AHS data. In order to produce estimates more representative of the entire census tract, I use data for all AHS renters.[14] In fact, one could argue that a prior based on all AHS households (including owner-occupied households) is most appropriate. Yet the overall AHS metro means are so far from most of the CSS tract means that I deem them unsuitable.

$\alpha_{prior}$ is set to the AHS weighted mean probability of a high rating multiplied by 4, and $\beta_{prior}$ is set to 4-$\alpha_{prior}$. This results in a prior Beta distribution with the same weighted mean as the AHS metro distribution, but a smaller sample size of 4 and a larger standard deviation.

Using the AHS number of weighted responses for the prior sample size would result in posterior distributions dominated by the AHS for most tracts. For a tract with the median CSS weighted responses close to 4, my choice of 4 for the prior sample size results in a posterior distribution where the AHS and CSS have approximately equal influence.

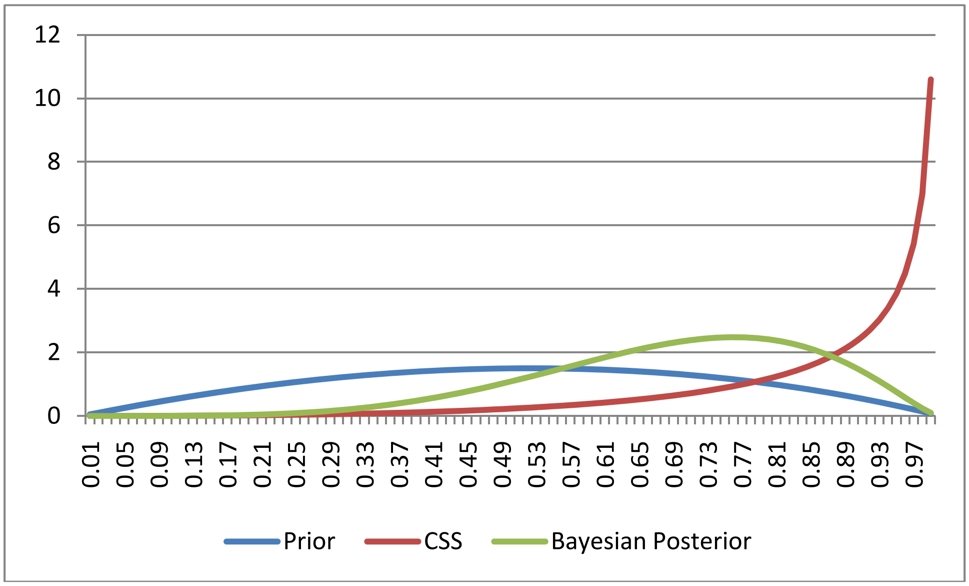
Exhibit 8 depicts the prior, CSS, and Bayesian posterior probability density functions for 1 randomly selected variable ($X_2$) in 1 tract (045031100) in 1 randomly selected metro area (Columbus). The tract was chosen because CSS weighted responses of 3.803 are closest to 4 (the median for all metro areas). The AHS-based metro prior distribution has 2.045 high ratings, 1.955 low ratings, and a mean probability of a high rating equal to (2.045/4) or .511. The prior standard deviation is .224. The CSS tract distribution is highly skewed, with 3.370 weighted high ratings, .433 weighted low ratings, and a mean probability of (3.370/3.803) or .866. The CSS tract standard deviation is .145.

The Bayesian posterior distribution is distributed Beta($\alpha^*$,$\beta^*$) with $\alpha^*$=2.045 + 3.370 =5.415, and $\beta^*$=.1.955 + .433 =2.388. The posterior mean probability equals 5.415/(5.415 + 2.338) or .694. Because the prior and CSS distribution have about the same sample size, the posterior mean is approximately a simple average of the prior and CSS means. The posterior standard deviation is .155.

As CSS weighted responses increase, the CSS data have greater influence on the posterior distribution. Exhibit 9 depicts the $X_2$ prior, CSS, and posterior probability density functions for Columbus area tract 041011520. The tract was chosen because it has 8.972 weighted CSS responses, which is closest to 9 (the 75th percentile for all metro areas). The metro-level prior distribution, described above, has mean .511. The CSS tract distribution has mean .329 and standard deviation .149. The posterior distribution has a mean probability of a high rating equal to .385. The CSS sample size is about 2.25 times that of the prior sample size of 4, thus the CSS has about 2.25 times the influence on the posterior distribution.

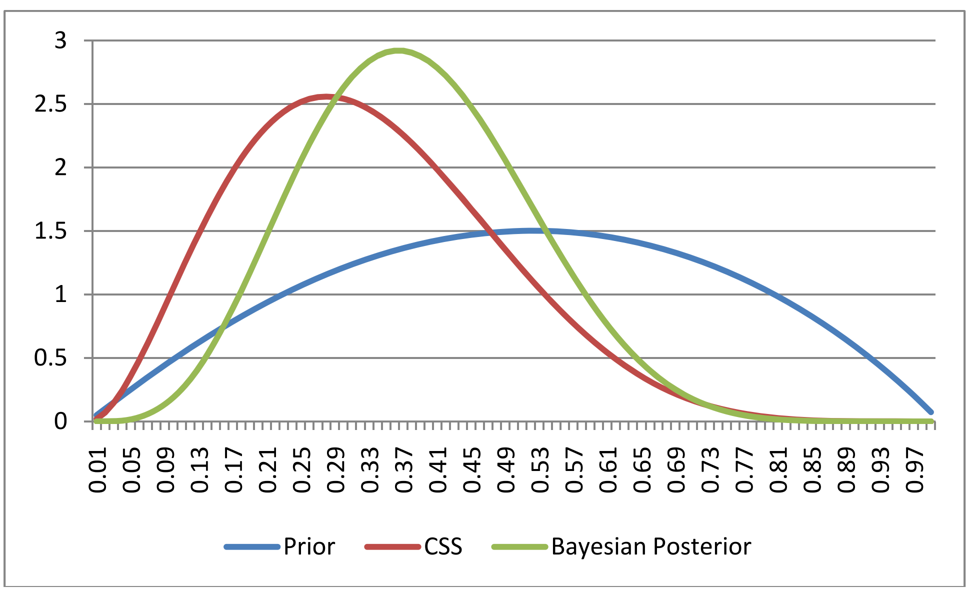
Exhibit 10 reports summary statistics for the Bayesian posterior means and standard deviations. Exhibit 11 depicts a histogram of CSS and Bayesian means for $X_4$(high neighborhood and home rating), which was randomly selected excluding $X_2$. The mean of the 3,749 Bayesian mean estimates for $X_4$ is .470, which is lower than the CSS mean of .527 reported in Exhibit 7. Compared to CSS estimates, The Bayesian estimates are much more normally distributed, with fewer tracts with very low or high means. The CSS estimates for $X_4$ have 692 tracts with mean=0, and 939 with mean=1. The Bayesian mean estimates, however, range from .028 to .983. And while the CSS standard deviations range from 0 to .443, the Bayesian standard deviations range from .010 to .221.

Exhibit 8: $X_2$ Prior, CSS, and Bayesian Posterior Probability Density Functions – Columbus Tract 045031100




Source: author's calculations using CSS and 2002 AHS data. $X_2$ is an indicator for a high home rating. Variable $X_2$ and the Columbus metro area were randomly chosen. The tract was chosen with CSS weighted responses closest to the median=4.

Exhibit 9: $X_2$ Prior, CSS, and Bayesian Posterior Probability Density Functions – Columbus Tract 041011520




Source: author's calculations using CSS and 2002 AHS data. $X_2$ is an indicator for a high home rating. Variable $X_2$ and the Columbus metro area were randomly chosen. The tract was chosen with CSS weighted responses closest to the $75^{th}$ percentile=9.

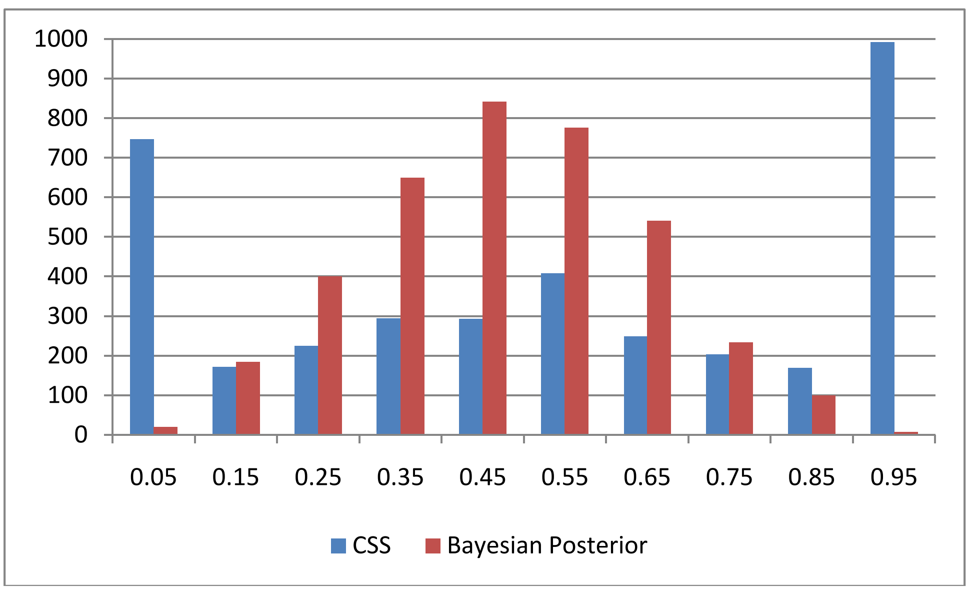
Exhibit 10: Summary Statistics for Bayesian Posterior Distributions

| | X1 (high neighborhood rating) | | X2 (high home rating) | | X3 (low crime perception) | |
|---|---|---|---|---|---|---|
| | Mean | Std | Mean | Std | Mean | Std |
| Minimum | 0.031 | 0.009 | 0.080 | 0.008 | 0.088 | 0.009 |
| Median | 0.577 | 0.156 | 0.598 | 0.156 | 0.764 | 0.133 |
| Mean | 0.558 | 0.152 | 0.581 | 0.152 | 0.726 | 0.132 |
| Maximum | 0.985 | 0.221 | 0.989 | 0.221 | 0.985 | 0.208 |

| | X4 (high neighborhood and home ratings) | | X5 (high neighborhood rating, low crime perception) | | X6 (high neighborhood and home ratings, low crime perception) | |
|---|---|---|---|---|---|---|
| | Mean | Std | Mean | Std | Mean | Std |
| Minimum | 0.028 | 0.010 | 0.023 | 0.011 | 0.021 | 0.011 |
| Median | 0.472 | 0.157 | 0.495 | 0.157 | 0.412 | 0.155 |
| Mean | 0.470 | 0.153 | 0.488 | 0.153 | 0.416 | 0.150 |
| Maximum | 0.983 | 0.221 | 0.973 | 0.221 | 0.970 | 0.221 |

N=3,749 tracts. Source: author's calculations using CSS and 2002 AHS data.

Exhibit 11: Histogram of CSS and Bayesian Mean Estimated Probabilities for $X_4$




N=3,749 tracts. Source: author's calculations using CSS and 2002 AHS data. $X_4$ is an indicator for a high neighborhood and home rating; it was randomly chosen excluding $X_2$.

## V) Data Validation

In this section I compare the CSS and Bayesian estimates with other tract level measures of neighborhood quality. These variables include median household income, percent of families living below the poverty line, and an indicator for 671 tracts qualifying for Low Income Housing Tax Credits (LIHTC).[15] Exhibit 12 reports summary statistics for these measures.

Exhibit 12: Summary Statistics for Auxiliary Neighborhood Quality Measures

| Variable | Minimum | Median | Mean | Std |
|---|---|---|---|---|
| Median income* | 7483.000 | 38946.000 | 40470.806 | 14473.096 |
| Poverty rate* | 0.280 | 12.140 | 15.029 | 10.814 |
| LIHTC indicator** | 0.000 | 0.000 | 0.179 | 0.383 |

N=3,749 tracts. Sources: *U.S. Census Bureau 2000 Census;
**http://www.huduser.org/datasets/lihtc.html .

Exhibit 13 reports Pearson correlation coefficients for the above auxiliary variables with the CSS and Bayesian estimated mean probabilities for $X_1$-$X_6$. All coefficients are significant at the .0001 level with the expected signs. Median income is positively related with neighborhood quality, while the poverty rate and the LIHTC indicator correlate negatively.

Exhibit 13: Pearson Correlation Coefficients

| | X1 (high neighborhood rating) | | X2 (high home rating) | | X3 (low crime perception) | |
|---|---|---|---|---|---|---|
| Auxiliary Variable | CSS mean | Bayesian mean | CSS mean | Bayesian mean | CSS mean | Bayesian mean |
| Poverty rate | -0.371 | -0.371 | -0.256 | -0.225 | -0.346 | -0.366 |
| Median income | 0.332 | 0.312 | 0.232 | 0.182 | 0.313 | 0.321 |
| LIHTC indicator | -0.263 | -0.270 | -0.169 | -0.162 | -0.238 | -0.260 |

| | X4 (high neighborhood and home ratings) | | X5 (high neighborhood rating, low crime perception) | | X6 (high neighborhood and home ratings, low crime perception) | |
|---|---|---|---|---|---|---|
| Auxiliary Variable | CSS mean | Bayesian mean | CSS mean | Bayesian mean | CSS mean | Bayesian mean |
| Poverty rate | -0.330 | -0.317 | -0.367 | -0.361 | -0.336 | -0.321 |
| Median income | 0.296 | 0.260 | 0.336 | 0.307 | 0.310 | 0.270 |
| LIHTC indicator | -0.233 | -0.237 | -0.256 | -0.262 | -0.235 | -0.240 |

N=3,749 tracts. All correlation coefficients are significant at the .0001 level. Source: author's calculations using U.S. Census Bureau, HUD LIHTC, AHS 2002, and CSS data.

For each neighborhood indicator, the CSS and Bayesian correlations with the auxiliary variables are very close. Of course, the Bayesian distributions are a weighted average of the prior and CSS distributions. As such, the CSS and Bayesian estimates are highly correlated. The Pearson correlation coefficient between the CSS and Bayesian means for $X_1$(high neighborhood rating) is .866. The Bayesian model is not intended to drastically change most of the CSS estimates; its purpose is to reduce outliers and make estimation possible for tracts with few CSS responses. Differences may be more apparent for tracts with larger differences between the CSS and Bayesian estimates.

Exhibit 14 reports Pearson correlation coefficients for tracts with an absolute percentage difference between CSS and Bayesian estimates at or above the median difference. Median differences range from about 14% for $X_3$ to about 35% for $X_6$. All of the correlations coefficients are significant at the .0001 level. For this subsample of tracts, 17 of the 18 Bayesian correlation coefficients are larger in absolute magnitude than their corresponding CSS coefficients.

Exhibit 14: Pearson Correlation Coefficients, Subsample of Tracts with Differences between CSS and Bayesian Estimates ≥ the Median

| | X1 (high neighborhood rating) | | X2 (high home rating) | | X3 (low crime perception) | |
|---|---|---|---|---|---|---|
| Auxiliary Variable | CSS mean | Bayesian mean | CSS mean | Bayesian mean | CSS mean | Bayesian mean |
| Poverty rate | -0.430 | -0.463 | -0.301 | -0.309 | -0.407 | -0.460 |
| Median income | 0.379 | 0.397 | 0.252 | 0.246 | 0.381 | 0.417 |
| LIHTC indicator | -0.320 | -0.339 | -0.206 | -0.214 | -0.268 | -0.302 |

| | X4 (high neighborhood and home ratings) | | X5 (high neighborhood rating, low crime perception) | | X6 (high neighborhood and home ratings, low crime perception) | |
|---|---|---|---|---|---|---|
| Auxiliary Variable | CSS mean | Bayesian mean | CSS mean | Bayesian mean | CSS mean | Bayesian mean |
| Poverty rate | -0.379 | -0.411 | -0.417 | -0.449 | -0.384 | -0.416 |
| Median income | 0.328 | 0.339 | 0.380 | 0.391 | 0.351 | 0.362 |
| LIHTC indicator | -0.282 | -0.304 | -0.300 | -0.318 | -0.280 | -0.306 |

All correlation coefficients are significant at the .0001 level. N=1,870 for $X_1$, 1,874 for $X_2$, $X_3$, and $X_5$, 1,869 for $X_4$, and 1,876 for $X_6$. Source: author's calculations using U.S. Census Bureau, HUD LIHTC, AHS 2002, and CSS data.

Exhibit 15 reports Pearson correlation coefficients for tracts with an absolute percentage difference between CSS and Bayesian estimates at or above the 66th percentile. Differences range from about 27% for $X_3$ to about 100% for $X_6$. All of the correlations coefficients are significant at the .0001 level. For this subsample of tracts, all Bayesian correlation coefficients are larger in absolute magnitude than their corresponding CSS coefficients. In addition, the differences between the CSS and Bayesian correlation coefficients are much larger for this subsample.

Exhibit 15: Pearson Correlation Coefficients, Subsample of Tracts with Differences between CSS and Bayesian Estimates ≥ the 66th Percentile

| | X1 (high neighborhood rating) | | X2 (high home rating) | | X3 (low crime perception) | |
|---|---|---|---|---|---|---|
| Auxiliary Variable | CSS mean | Bayesian mean | CSS mean | Bayesian mean | CSS mean | Bayesian mean |
| Poverty rate | -0.422 | -0.492 | -0.348 | -0.402 | -0.382 | -0.468 |
| Median income | 0.398 | 0.448 | 0.308 | 0.343 | 0.408 | 0.468 |
| LIHTC indicator | -0.294 | -0.340 | -0.217 | -0.254 | -0.244 | -0.290 |

| | X4 (high neighborhood and home ratings) | | X5 (high neighborhood rating, low crime perception) | | X6 (high neighborhood and home ratings, low crime perception) | |
|---|---|---|---|---|---|---|
| Auxiliary Variable | CSS mean | Bayesian mean | CSS mean | Bayesian mean | CSS mean | Bayesian mean |
| Poverty rate | -0.346 | -0.446 | -0.373 | -0.465 | -0.345 | -0.432 |
| Median income | 0.335 | 0.410 | 0.382 | 0.441 | 0.349 | 0.413 |
| LIHTC indicator | -0.223 | -0.300 | -0.250 | -0.312 | -0.211 | -0.285 |

All correlation coefficients are significant at the .0001 level. N=931 for $X_1$, 938 for $X_2$, $X_3$, and $X_5$, 941 for $X_4$, and 1,031 for $X_6$. Source: author's calculations using U.S. Census Bureau, HUD LIHTC, AHS 2002, and CSS data.

## VI) Conclusion

While neighborhood quality is important for public policy, it is also difficult to quantify. This study measures neighborhood quality using data from two sources: the 2002 American Housing Survey (AHS), and HUD's Customer Satisfaction Survey (CSS) of Section 8 Housing Choice Voucher Program (HCVP) households.

While the AHS and CSS contain related questions, differences in survey methods and question wording make direct comparison of the two surveys difficult. However, Bayesian methods are flexible enough to use information from related questions from both surveys.

I examine survey responses in 13 metro areas regarding neighborhood quality, home quality, and crime perceptions. Tract level Bayesian estimates are computed using AHS metro level data and CSS census tract data.

Compared to estimates solely based on CSS data, the Bayesian estimates have fewer outliers. Bayesian analysis also allows for estimation for tracts with lower sample sizes than would be practical using only CSS data.

I Compare the CSS and Bayesian estimates with other measures of neighborhood quality, such as poverty rates, median income, and indicators for tracts receiving Low Income Housing Tax Credits. The Bayesian estimates tend to correlate more strongly with other neighborhood quality indicators. The CSS and Bayesian indicators are highly correlated, and the differences are more apparent for tracts with larger differences between the CSS and Bayesian estimates.

References

Austin, D. Mark, L. Allen Furr, and Michael Spine. 2002 "The effects of neighborhood conditions on perceptions of safety" Journal of Criminal Justice Vol 30(5), September-October, Pages 417-427.

Bishaw ,Alemayehu and Sharon Stern. 2006 (June 15) "Evaluation of Poverty Estimates: A Comparison of the American Community Survey and the Current Population Survey" U.S. Census Bureau http://www.census.gov/hhes/www/poverty/acs_cpspovcompreport.pdf .

Buron, Larry, Bulbul Kaul, and Rhiannon Patterson. 2003. *Quality of Housing Choice Voucher Housing.* Report prepared for the U.S. Department of Housing and Urban Development.

Buron, Larry and Satyendra Patrabansh. 2008 "Are Census Variables Highly Correlated With Housing Choice Voucher Holders' Perception of the Quality of Their Neighborhoods?," *Cityscape A Journal of Policy Development and Research Vol 10(1), pp. 157-184.*

Cahill, Meagan Elizabeth, 2006. Geographies of Urban Crime in Nashville, Tennessee, Portland, Oregon, and Tucson, Arizona, 1998-2002, Ann Arbor, MI: Inter-university Consortium for Political and Social Research, August http://www.icpsr.umich.edu/cocoon/NACJD/STUDY/04547.xml .

Chapman, David W and; Lombard, John R. 2006. Determinants of Neighborhood Satisfaction in Fee-Based Gated and Nongated Communities. *Urban Affairs Review* 41(6), July, pp. 769-99. http://www.sagepub.com/journalsProdDesc.nav?prodId=Journal200784

Craig, St. John and Frieda Clark. 1984. "Racial differences in dimensions of neighborhood satisfaction," Social Indicators Research 15(1): 43-60.

Deller, Steven C. and Thomas Ottem, 1991 (July). "Crime and Quality of Life in Wisconsin Counties," Department of Agricultural and Applied Economics, University of Wisconsin-Madison, Staff Paper Series No. 442. http://www.aae.wisc.edu/pubs/sps/pdf/stpap442.pdf .

Dilulio, John J., Jr. 1994. "The question of black crime." *Public Interest* (117), Fall, p. 3.

Dubin, Robin A., 1992 (September). "Spatial autocorrelation and neighborhood quality," *Regional Science and Urban Economics*, Elsevier, vol. 22(3), pages 433-452.

Gelman, Andrew et al. *Bayesian Data Analysis: 2nd Edition*, 2004 Chapman and Hall/CRC, London.

Gray, Robert W., Barbara A. Haley, and Brent D. Mast. 2009 (May). "Tell Us About Your Home: Three Years of Surveying Housing Quality and Satisfaction in the Section 8 Housing Choice Voucher Program" Unpublished manuscript. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Hipp, John. Resident Perceptions of Crime: How Similar are They to Official Crime Rates? Center for Economic Studies, U.S. Census Bureau, Working Papers. December, 2007 http://www.ces.census.gov/index.php/ces/cespapers?detail_key=101783 .

Lee, Peter M. 2004. *Bayesian Statistics: An Introduction, 3rd Edition*. Hodder Arnold, London.

Mast, Brent D. May 2009a (May) "Bayesian Hierarchical Model Estimates of Local Crime Perceptions," working paper. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research. http://www.huduser.org/publications/pmrd/pmrd_workpap1.html

Mast, Brent D. 2009b (July). "Measuring Housing Quality in the Housing Choice Voucher Program with Customer Satisfaction Survey Data," *Cityscape: A Journal of Policy Development and Research* Vol. 11(2). http://www.huduser.org/periodicals/cityscpe/vol11num2/index.html

Newman, Sandra J.and Schnare, Ann B.1993. "Last in Line: Housing Assistance for Households with Children," *Housing Policy Debate*. 4(3), pp. 417-55. http://www.fanniemaefoundation.org/programs/hpd.shtml

Sedgley, Norman H., Nancy H. Williams and Frederick W. Derrick, 2008 (June). "The effect of educational test scores on house prices in a model with spatial dependence," *Journal of Housing Economics* Vol. 17(2).

U.S. Department of Housing and Urban Development, 2008. Housing Choice Voucher Handbook, accessed September 2009, http://www.hud.gov/offices/pih/programs/hcv/forms/guidebook.cfm .

Notes

---

[1] AHS data and information are available at http://www.huduser.org/datasets/ahs.html .

[2] See Mast (2009b) and Gray, Haley, and Mast (2009) for more CSS information.

[3] See Gray, Haley, and Mast (2009) and Buron, Kaul, and Patterson (2003) for discussions of differences between the AHS and CSS.

[4] "Bayesians view statistical inference as a problem in *belief dynamics*, or use of evidence about a phenomenon to revise and update our knowledge about it" http://volgenau.gmu.edu/~klaskey/SYST664/SYST664.html. See Lee (2004) for an introduction to Bayesian methods.

[5] Designed to replace the decennial Census long form, the American Community Survey (ACS) will soon provide tract level estimates (averaged over multiple years).

[6] Cahill (2006) data and information are available at http://www.icpsr.umich.edu/cocoon/NACJD/STUDY/04547.xml .

[7] http://spotcrime.com/wa/seattle .

[8] CSS data were matched to the AHS by county for counties in the 13 metro areas according to OMB June 1999 definitions. For information on metropolitan area definitions, see http://www.census.gov/population/www/metroareas/metrodef.html.

[9] The metro AHS is stratified by metro area, with weights summing to total households. The CSS is stratified by PHA and year, with weights summing to HCVP households in all sampled PHAs in a given year. Only a tiny fraction of PHAs were excluded. See supra notes 1 and 2 for more information on the AHS and CSS survey designs, respectively.

[10] There are limited combinations of distributions for data and parameters, referred to as conjugate pairs, with analytic solutions for Bayesian posterior distributions. While there is a conjugate model for multinomial categorical data, it does not account for ordering of the categories. Therefore I employ a binomial-beta conjugate model, where the household neighborhood indicators are binomial and the probability of a high rating follows a beta distribution. For a Bayesian analysis of AHS and CSS data with a normal-normal conjugate model, see Mast (2009a).

[11] I consider a CSS tract sample proportion to be normally distributed if weighted responses are at least 30 and each binary category has at least 5 weighted responses.

[12] Let $W_i$ be the original survey weights with n responses summing to population, and let $W_i^*$ be the adjusted weight summing to responses: $W_i^* = nW_i/\sum_{i=1}^{n} Wi$ .

[13] For an accessible introduction to Bayesian Hierarchical Models, see http://volgenau.gmu.edu/~klaskey/SYST664/Bayes_Unit5.pdf .

[14] I also produced Bayesian posterior estimates using metro priors based on the AHS HCVP sample. These estimates had lower correlation with auxiliary variables compared to estimates with priors based on all renters (results available upon request).

[15] LIHTC data indicators are for tracts qualifying in any year between 2000 and 2003. Original data were for qualifying tracts based on 1990 geography. For this study, I constructed qualifying tracts based on 2000 geography. For tracts that changed, I assumed a tract with 2000 geography qualified if it included any part of a tract qualified based on 1990 geography.