U.S. Department of Housing and Urban Development
Office of Policy Development and Research

American Housing Suvey

Combining the American Housing Survey and the American Community Survey to Produce Information Useful in Public Emergency Situations:

An Exploratory Analysis

Visit PD&R's Web Site
www.huduser.org
to find this report and others sponsored by
HUD's Office of Policy Development and Research (PD&R).
Other services of HUD USER, PD&R's Research Information Service,
include listservs;special interest, bimonthly publications (best practices, significant studies from other sources); access to public use databases; hotline 1-800-245-2691 for help accessing the information you need.

American Housing Survey

Combining the American Housing Survey and the American Community Survey to Produce Information Useful in Public Emergency Situations: An Exploratory Analysis

Prepared for:
U.S. Department of Housing & Urban Development
Office of Policy Development & Research

Prepared by:
Frederick J. Eggers
Econometrica, Inc.
Bethesda, Maryland

Contract No. C-CHI-00976
Order No. CHI-T0001
Project No. 1026-001

December 2009

Acknowledgements

The author is indebted to several persons for their help on this challenging task. Dr. Fouad Moumen of Econometrica, Inc., assisted in developing the independent and dependent variables used for all the regressions and carried out many of these analyses. Dr. Fred Bellemore and Mr. Chris Holleyman of Econometrica, Inc., performed the fractional logit analyses using STATA© and provided advice and other assistance. Dr. Albert Lee of Summit Consulting provided highly useful critiques of the regression analyses and suggested the use of fractional logit. Dr. David A. Vandenbroucke of the Office of Policy Development and Research at HUD helped identify policy-useful dependent variables and offered other important guidance throughout the project.

Table of Contents

Acknowledgements*... *v

Executive Summary*.. *vi

I. The Analytical Framework... 1

Small Area Estimation... 2
National vs. Metropolitan AHS 2
Unit Data vs. Data Grouped by Metropolitan Area 3

II. Emergency-Relevant AHS Variables 3

III. Data Issues 8

Metropolitan Geography... 8
Small Sample Sizes 10
Choice of Statistical Technique 12

IV. Covariates for Modeling 13

ACS Approximations of the AHS Condition.. 13
Covariates to Distinguish Among Metropolitan Areas on Non-Housing Factors 15
Covariates Related to Local Housing Markets and Economic Conditions.............. 16

V. Statistical Models.. 17

VI. Conclusions... 21

Appendix A: ACS and AHS Metropolitan Areas A-1

Appendix B: Correlation Matrix for Independent Variables.......................... B-1

Appendix C: Fractional Logit Estimates for the Other Dependent Variables C-1

List of Tables

Table 1: Types of Information Published by the AHS and the ACS................................ 4

Table 2: Correlations among the AHS Condition Variables... 9

Table 3: Sample Sizes and Descriptive Statistics for the AHS Condition Variables 11

Table 4: Correlations among AHS Condition Variables and their ACS Approximations15

Table 5: Chi-squared and R^2 Statistics.. 17

Table 6: Regression Results 19

Table 7: Predicted Values of PCTHSHLDNEED and PCTRENTINS 20

Executive Summary

This research project explores the possibility of using small area statistical techniques to generate for areas not covered by the American Housing Survey (AHS) information that would be useful in preparing for or responding to a disaster and that is available only in the AHS.

The study experienced difficulty in identifying variables that would be of *obvious* help in disasters situations. It looks at nine conditions that might be useful to measure:

- the percent of households with needs—households that (a) are poor, (b) have an elderly householder or spouse, (c) are single-parent households, or (d) have a householder or spouse who is a recent immigrant
- the proportion of occupied units with severe physical problems
- the proportion of occupied units with either severe or moderate physical problems
- the proportion of occupied units that have severe physical problems and are occupied by households with needs
- the proportion of occupied units that have severe or moderate physical problems and are occupied by households with needs
- the proportion of occupied units built prior to 1940 and are occupied by households in need
- the proportion of occupied units that are mobile homes and are occupied by households in need
- the proportion of owner-occupied units with homeowners' insurance
- the proportion of renter-occupied units with renters' insurance.

The research chose to use fractional logit to estimate these conditions using variables measured for all metropolitan areas by the annual American Community Survey (ACS). Three types of independent variables were used: ACS approximations of the AHS-measured conditions, covariates to distinguish among metropolitan areas on non-housing factors, and covariates related to local housing markets and economic conditions.

The results are only mildly encouraging. Of the nine equations estimated, only two were statistically meaningful using the Chi-squared test. However, the R^2s for these two equations indicate that using the predicted values may be a worthwhile improvement over assuming that all metropolitan areas have the same values, that is, using the sample mean. Out-of-sample predictions using these two equations suggest that conditions may vary substantially across metropolitan areas.

Equations with better statistical properties might result from using household-level data rather than metropolitan-level data. This would involve using microdata from the AHS to estimate the equations and microdata from the ACS public use files to use the equations for predictive purposes.

The next major disaster may very well indicate what AHS information is crucial. Unfortunately, if such an event were to occur, there may not be time to carry out the desired microdata analysis. If so, the techniques explored in this paper might be a valuable second-best approach.

Combining the American Housing Survey and the American Community Survey to Produce Information Useful in Public Emergency Situations: An Exploratory Analysis

In October 2008, Econometrica, Inc., and its partner, ICF International, entered into a contract with the Department of Housing and Urban Development (HUD) to support the American Housing Survey (AHS). Task C of that contract required the Econometrica team to use the information from the 2007 AHS surveys of seven metropolitan areas to produce information for those areas that might be useful to federal, state, or local agencies in preparing for or responding to natural disasters or other emergencies. For Task D, Econometrica suggested exploring the possibility of using small area statistical techniques to generate information available only in the AHS to areas not covered by the AHS. ICF International performed the analysis under Task C; Econometrica carried out the work under Task D.

This paper reports in six sections what was learned in Task D:

- Section I translates the Task D idea into research terms.
- Section II looks for information available only in the AHS that might be useful in emergency situations.
- Section III discusses relevant data issues and statistical issues.
- Section IV identifies the covariates to be used in regression analysis to estimate AHS variables using American Community Survey information.
- Section V reports the model results.
- Section VI provides brief concluding observations.

I. The Analytical Framework

A simple concept underlies Task D:

> At present, the American Community Survey (ACS) produces detailed information annually on people and housing for all geographical units with 20,000 or more persons.[1] Beginning in late 2010, the ACS will provide the same information for all places and for census tracts and census blocks as well.[2] Federal, state, and local agencies preparing for

---

[1] For places with more than 65,000 persons, the ACS publishes information using data collected in the preceding year. For places with more than 20,000 persons (including places with more than 65,000 persons), the ACS publishes information using data collected in the preceding 3 years.

[2] For places with fewer than 20,000 persons and for census tracts and blocks, the ACS will publish information using data collected in the preceding 5 years. Information based on the same time period will also be published for all places, including those with populations greater than 20,000. Once the ACS is fully implemented in 2010, the statistics available annually from the ACS by size of place are as follows: For places with populations greater than 65,000, the Census Bureau will release three different data sets, one based on data collected in the preceding year, one based on data collected in the preceding 3 years, and one based on data collected in the preceding 5 years. Places with populations greater than 20,000 but less than 65,000 will have two data sets, one based on data collected

or responding to natural disasters or other emergencies should find ACS information very valuable. The AHS contains more detailed information on housing units and the households living in those units than the ACS. This additional information might also be valuable to those dealing with emergencies. However, AHS data are generally not useful in emergency situations because AHS data are available only at the national or regional level. Task D explores whether statistical models using information from the ACS can produce satisfactory estimates of AHS variables for metropolitan areas for which AHS data are not available.

Small Area Estimation

"Small area estimation" is the term applied to a set of statistical methods that have been developed in recent years to solve problems such as the one posed in the preceding paragraph, namely, to produces estimates for "areas" or "domains" for which the sample sizes available in a "national" sample are too small for direct estimation. An example of small area estimation is the Bureau of Labor Statistics (BLS) Local Area Unemployment Statistics (LAUS) series that uses the national Current Population Survey to generate estimates of labor force participation, employment, and unemployment for small areas. Because of the considerable work that went into the creation and testing of the statistical models used in LAUS, the BLS is confident enough in their reliability to release LAUS estimates on a monthly basis for jurisdictions with populations as small as 10,000. The goal of this task is considerably more modest. HUD does not intend to publish any estimates from the models created in this task. The goal is simply to test the feasibility of developing techniques that could be applied quickly in an emergency situation to generate information useful to those responding to the emergency.

A variety of statistical tools are used to solve small area estimation problems. The author uses one of the most frequently employed techniques, regression analysis, but the regression equations used here are considerably simpler than the models used to create publishable small area statistics. Two reasons explain the choice of simpler models: the limited scope of the task and the objective of creating techniques that could be replicated quickly and simply, even with different dependent variables or different levels of geography.

Translating this concept into a framework suitable for statistical modeling involves answering the two questions discussed below.

National vs. Metropolitan AHS

Should the analysis be based on the national AHS or on metropolitan AHS surveys?

The AHS surveys 21 metropolitan areas on a once-every-six-year schedule. The public use files (PUFs) contain data at the household level with location identified down to "zones," which are units of geography containing no fewer than 100,000 people, created especially for the AHS.[3] It

in the preceding 3 years and one based on data collected in the preceding 5 years. Places with populations less than 20,000 will have only one data set based on data collected in the preceding 5 years.

[3] "Zones" are not available for the New York, Los Angeles, Chicago, Detroit, and Philadelphia metropolitan areas and for the group of metropolitan areas surveyed in the AHS as "Northern New Jersey."

might be useful to project AHS-type information down to smaller levels of geography such as census tracts or blocks. The author chose not to pursue this option for two reasons. First, the author could not do this without access to information on the location of AHS sample units that the Census Bureau does not make available to the public. While it would be possible to petition the Census Bureau for access to this information, the process of obtaining the needed clearances seemed to be beyond the timeframe of this Task. Second, ACS data are not yet available at the census tract level. While the author could use 2000 decennial census data, this information would be 9 years old. For these reasons, the author chose to use the national AHS data.

Unit Data vs. Data Grouped by Metropolitan Area

Should the analysis use household-level data or metropolitan-level data?

The AHS publishes information at the national and regional levels. In addition, the Census Bureau makes all of the AHS data available in a public use file (PUF) that identifies 144 metropolitan areas.[4] The ACS publishes data for various levels of geography, including most metropolitan areas. The Census Bureau also releases a PUF for the ACS containing roughly a million cases at the level of Public Use Microdata Area (PUMA), which are units of geography containing no fewer than 100,000 people and created especially for the ACS.

The first option would involve using the AHS PUF data to model relationships at the housing-unit level. Then one could apply the AHS models to PUMA geography using microdata from the ACS PUF. The second option would involve using the AHS PUF to group AHS data by metropolitan area, and then build models that use AHS information at the metropolitan level as dependent variables and ACS published information at the metropolitan level as independent variables. Then one could apply these models to other metropolitan areas and conceivably to lower levels of geography for which ACS published information could be used for independent variables.

The author chose the second option because it could be implemented quickly if an emergency arose. A model created using published ACS data could be applied swiftly to produce an estimate of AHS variables for any metropolitan area. As explained in the next section, there was little consensus about what AHS variables would be most useful to those responding to an emergency situation. The simple models based on area-level data developed in this task could be easily replicated using different AHS variables as dependent variables. Finally, if the estimated models appear to be robust and if the exigency were great, one could even apply the model to lower levels of geography quickly using published ACS data.

II. Emergency-Relevant AHS Variables

While the ACS contains a considerable amount of useful information on housing, the AHS provides substantially more information on housing units, including detailed information on amenities, housing costs, and the condition of housing units. The following table is excerpted

[4] The number of sampled households in each of the identified metropolitan areas ranges from 8 to 1,274. The PUF contains information on over 35,000 households throughout the United States.

from *Comparison of Housing Information from the American Housing Survey and the American Community Survey*, a report prepared for HUD by Econometrica in 2007.[5]

Table 1: Types of Information Published by the AHS and the ACS

The italic entries represent housing characteristics available in published tables from the AHS that are not available in published tables from the ACS.

	American Housing Survey (national survey)	American Community Survey
Status of unit	• Seasonal, year-round, occupied/vacant, reason vacant • Owner/renter	• Seasonal, occupied/vacant, reason vacant • Owner/renter
Structure characteristics	• Units in structure • Year built • *Stories in structure* • *Lot size (single unit and mobile homes only)*	• Units in structure • Year built
Unit characteristics	• Number of rooms • Number of bedrooms • *Number of bathrooms* • *Square footage* • *Heat equipment* and heating fuel • *Source of water and sewage disposal*	• Number of rooms • Number of bedrooms • Heating fuel
Unit quality	• With all kitchen facilities—*by specific facility* • With all plumbing facilities—*by specific facility* • *Selected amenities, includes:* o Telephone service o *Air conditioning* o *Fireplace* o *Washing machine* o *Clothes dryer* o *Dishwasher* o *Disposal* o *Deck* o *Garage or car port* • *Selected deficiencies, includes:* o *Selected structural deficiencies, e.g., leaks or holes in floors or ceilings* o *Selected plumbing problems* o *Selected electrical problems* o *Selected upkeep problems* • *Summary measure of severe and moderate physical problems* • *Occupants rating of unit on scale of 1 to 10*	• With all kitchen facilities • With all plumbing facilities • Telephone service
Neighborhood characteristics	• *Type and age of housing in neighborhood*	• No national-level tables • Beginning in 2010, researchers

[5] This report is available at http://www.huduser.org/intercept.asp?loc=/publications/pdf/comparison_hsg.pdf.

	American Housing Survey (national survey)	American Community Survey

and quality	• *Adequacy of schools, shopping, and public transportation*
	• *Problems with:*
	  o *Crime*
	  o *Traffic noise*
	  o *Odors*
	  o *Abandoned buildings*
	  o *Litter*
	• *Occupants rating of neighborhood on scale of 1 to 10*
		will be able to obtain for individual census tracts information on type and age of housing in census tract, housing costs in census tract, home values in census tract, and household characteristics in census tract

Renter housing costs	• Contract rent • *Utilities and other housing costs* • Gross rent—contract rent plus utilities and other housing costs • Ratio of gross rent to income • *Whether unit receives housing assistance*	• Contract rent • Gross rent • Ratio of gross rent to income
Owner housing costs	• *Utilities and other housing costs* • Real estates taxes • *Mortgage payments* • Monthly housing costs • Ratio of monthly housing costs to income	• Real estates taxes • Monthly housing costs • Ratio of monthly housing costs to income
Value of owner-occupied housing	• *Purchase price* • *Source of downpayment* • Owner estimate of value • Value by income • Ratio of value to income • Price asked for vacant units	• Owner estimate of value • Value by income • Ratio of value to income • Price asked for vacant units
Mortgage	• With or without prime mortgage • *How obtained* • *Year originated* • *Government insurance* • *Payment type, e.g., fixed-rate* • *Mortgage payment* • *Interest rate* • *Remaining years* • *Outstanding principal amount* • *Current loan as percent of value* • *Second mortgage* • *Home equity line of credit*	• With or without prime mortgage

Despite all the extra information available in the AHS, the author found it difficult to identify information unique to the AHS that would *obviously* be useful in emergency situations. The author used the literature review produced for Task C, discussions with HUD analysts, previous work by the author on permanent losses,[6] examination of information produced in the past by

[6] *The Destruction of Housing Capital: A Preliminary Exploration into Demolitions and Disasters*. This 2004 ICF-Econometrica report was produced for HUD and can be found at http://www.huduser.org/datasets/ahs/ahsReports.html#1.

HUD using the AHS, and considerable introspection and still failed to find variables that would be unquestionably useful. As frequently happens, it is difficult to anticipate whether something might be useful in the abstract. In discussing this difficulty, David Chase noted that, in the early days of geocoding data, it was difficult to get policymakers and even analysts to specify what maps would be useful until the policymakers and analysts became familiar with what mapping could do.

The author selected two measures of deficient housing that are unique to the AHS:

- PCTSEVPROB: the proportion of occupied units with severe physical problems.[7]

  Housing with severe physical problems may be more susceptible to damage or, if damaged, may be deemed unfit or not valuable enough to repair. However, analysis using data on individual units in *The Destruction of Housing Capital* did not find an association between units with severe physical problems and permanent losses through demolition or disasters.

- PCTSEVMODPROB: the proportion of occupied units with either severe or moderate physical problems.

  Having either moderate or severe physical problems was related statistically to permanent losses through demolition or disasters in *The Destruction of Housing Capital*. Units with problems of these types may be more susceptible to damage or, if damaged, may be deemed unfit or not valuable enough to repair. In either case, provision for new housing must be made.

The cited research on permanent losses also found statistical relationships of varying strength at the individual unit level between permanent losses through demolition or disasters and the following AHS variables:[8]

- Having been built prior to 1920 – a consistently strong statistical relationship
- Having been built between 1920 and 1939 or having been built between 1940 and 1959 – consistent statistical relationships of varying strength
- Midwest location – consistent statistical relationship of varying strength
- Renter-occupancy – consistently strong relationship
- Vacancy – consistently strong relationship.

The ACS provides information on all these variables, and therefore the author does not model them.

---

[7] For definitions of severe and moderate physical problems, see pages 1,043 and 1,044 of the AHS Codebook at http://www.huduser.org/intercept.asp?loc=/Datasets/ahs/AHS_Codebook.pdf. Changes to the questionnaire in 2007 eliminated the questions needed to assess the adequacy of hallways.

[8] *The Destruction of Housing Capital* also found a consistent and strong statistical relationship between an occupant's rating of his or her housing unit and permanent losses through demolition or disaster. This variable (HOWH) is not available in the ACS, but its granularity (it is scored on a 1-10 scale) makes it difficult to create an appropriate summary variable for an entire metropolitan area.

The author created a measure of households in need:

- PCTNEEDHSHLD: the percent of households that (a) are poor (below poverty-level income), (b) have an elderly householder or spouse, (c) are single-parent households, or (d) have a householder or spouse who is a recent immigrant (came to this country in 2002 or later).

The information on households with needs could be obtained from the ACS PUF, so the AHS is not a unique source.[9] The author used the AHS to identify households in need so that the author could create variables that link housing problems and household needs. However, the author decided to estimate this variable as well, because it might be useful to have an equation that estimated this percentage without having to use the ACS PUF.

The author linked being a household in need with specific housing features in the following four variables:

- PCTSEVPROBNEED: the proportion of occupied units that have severe physical problems and are occupied by households with needs.

- PCTSEVMODPROBNEED: the proportion of occupied units that have severe or moderate physical problems and are occupied by households with needs.

- PCTOLDHSNEED: the proportion of occupied units built prior to 1940 and are occupied by households in need.

- PCTMHNEED: the proportion of occupied units that are mobile homes and are occupied by households in need.

The author focuses on older housing units because of the relationship found in *The Destruction of Housing Capital* between the age of a unit and permanent losses through demolition or disaster. The author focuses on mobile homes because of their susceptibility to wind damage. Because the ACS contains information on both year built and type of structure, the author estimates only the intersection of these structural characteristics and household needs.

HUD suggested including a variable that identifies whether the household has homeowners' or renters' insurance. The percentage of owners with homeowners' insurance is substantially greater than the percentage of renters with renters' insurance.[10] Therefore, the author decided to estimate two variables:

[9] Table B17017 in the published ACS data contains counts of households by all of these conditions except being a recent immigrant; the information on being elderly in Table B17017 refers to the householder only. Separately, the ACS also contains information on household members with disabilities. The AHS will have this information beginning in 2009.
[10] Among the 118 metropolitan areas for which we calculate the percentages, 25 percent of renters had renters' insurance and 95 percent of homeowners had homeowners' insurance.

- PCTOWNINS: the proportion of owner-occupied units with homeowners' insurance.

- PCTRENTINS: the proportion of renter-occupied units with renters' insurance.

Table 2 shows that, except for PCTSEVMODPROB and PCTSEVMODPROBNEED, the correlations between these seven AHS condition variables are relatively small.

III. Data Issues

The author confronted three data issues: how to reconcile the metropolitan areas reported in the AHS with those reported in the ACS, what to do about metropolitan areas where the AHS sample was small, and what statistical technique to use for the modeling.

Metropolitan Geography

The 2007 AHS PUF identifies 144 separate metropolitan areas using the variable SMSA and identifies an additional 3 geographical areas that are a combination of some of the 144 metropolitan areas. The ACS publishes 2007 data on 310 metropolitan areas. Matching the AHS metropolitan areas with those found in the ACS involves taking into account changing definitions of metropolitan areas and the availability of published ACS data. Appendix A provides a list of the 118 metropolitan areas used in this analysis. These are areas for which the author had both data from the 2007 national AHS and published data from the 2007 ACS.

For 103 of the 118 metropolitan areas, there was a one-to-one match between the areas identified in the AHS and those for which 2007 ACS data are published. There may, however, be differences in the boundaries of an AHS metropolitan area and the ACS metropolitan area with the same name. The author cannot correct for boundary differences.

The remaining 13 metropolitan areas involved fitting together multiple AHS metropolitan areas or using ACS combined metropolitan areas.

- In four cases, the author had to use combined statistical areas when obtaining the published ACS data.
    - The AHS reports Johnson City-Kingsport-Bristol, TN-VA as one metropolitan area; the ACS reports them as separate areas and also publishes data for the two areas combined.
    - The AHS reports Raleigh-Durham as one metropolitan area; the ACS reports them as separate areas and also publishes data for the two areas combined.
    - Both the AHS and ACS recognize Orlando and Daytona Beach as separate metropolitan areas, but the ACS only publishes data for Orlando and for the combination of the two metropolitan areas.
    - Both the AHS and ACS recognize Greenville, SC and Spartanburg, SC as separate metropolitan areas, but the ACS only publishes data for Greenville and for the combination of the two metropolitan areas.

Table 2: Correlations among the AHS Condition Variables

	PCTSEVPROB	PCTSEVMODPROB	PCTNEEDHSHLD	PCTSEVPROBNEED	PCTSEVMODPROBNEED	PCTOLDHSNEED	PCTMHNEED	PCTOWNINC	PCTRENTINS
PCTSEVPROB	1								
PCTSEVMODPROB	0.5182	1							
PCTNEEDHSHLD	-0.0557	-0.0161	1						
PCTSEVPROBNEED	0.6114	0.2742	-0.0272	1					
PCTSEVMODPROBNEED	0.3545	0.8188	0.2161	0.3388	1				
PCTOLDHSNEED	-0.1143	-0.0711	0.3735	0.0240	0.1325	1			
PCTMHNEED	0.0700	-0.0189	0.2350	-0.0183	-0.1089	-0.1938	1		
PCTOWNINS	0.0276	-0.4232	-0.2484	0.0446	-0.4092	-0.0432	-0.0766	1	
PCTRENTINS	-0.1213	-0.0653	-0.0923	-0.0866	-0.0572	0.0436	-0.1337	0.0477	1

- The ACS New York City metropolitan area includes all of Long Island and most of northern New Jersey. The ACS definition encompasses seven metropolitan areas which the AHS identifies separately. In addition, the AHS separately identifies data that fall into one of the five northern New Jersey metropolitan areas without specifying which metropolitan area, and also data that fall into the New York City or Nassau-Suffolk metropolitan areas without identifying which area.[11]

- The ACS Chicago metropolitan area has boundaries that include five AHS metropolitan areas, including Gary, IN. In addition, the AHS separately identifies data that fall into one of four metropolitan areas included in the greater Chicago area, without specifying which metropolitan area.

- The ACS metropolitan area includes two or more AHS areas in the following nine cases: Boston (3), Bridgeport-Stamford (2), Dallas-Fort Worth (2), Los Angeles-Long Beach (2 – the second being Santa Ana), Miami-Ft. Lauderdale (3), Pittsburg (2), St. Louis (3), San Francisco-Oakland (2), and Seattle-Tacoma (2).

Appendix A lists all 118 ACS areas used in this analysis and the AHS metropolitan areas that are associated with them.

Small Sample Sizes

In the models developed in this research project, the author used AHS data only for constructing the dependent variables and used ACS data for all the independent variables.[12] For many of the metropolitan areas, the AHS sample sizes are small, and therefore estimates based on these samples will be imprecise and granular. For example, if there are only 20 AHS sample cases in a metropolitan area, then all percentages will be multiples of 5. Normally, imprecision in estimating the dependent variable does not introduce bias into the coefficient estimates, but bias may result here because the independent variables are percentages and most of the observed percentages are zero or close to zero. Small sample sizes result in vectors of values for the dependent variables that appear to contain too many zeros.

Table 3 calls attention to the fact that, except for having homeowners' insurance, most of the AHS conditions occur infrequently; this is particularly true when a physical condition is combined with household need. A unit with a severe physical problem is a condition that occurred on average in only 1.6 percent of the units in the 118 metropolitan areas to be used in the regression for PCTSEVPROB.[13] A unit with a severe physical problem and also occupied by a household with one or more needs occurs in only 0.5 percent of the units in the 118 metropolitan areas to be used in the regression for PCTSEVPROBNEED. The coefficient of

[11] The AHS data on northern New Jersey (SMSA=9993) include some cases in the Trenton metropolitan area. The author elected to treat Trenton as a separate metropolitan area and to ignore the erroneous inclusion in the New York metropolitan area of any SMSA= 9993 cases that are in the Trenton metropolitan area.
[12] AHS data were used for REGION for convenience; the AHS is not a necessary source for that information.
[13] The means reported in the table are unweighted means of the percentages observed in the different metropolitan areas.

Table 3: Sample Sizes and Descriptive Statistics for the AHS Condition Variables

AHS CONDITION	PCTSEVPROB	PCTSEVMODPROB	PCTNEEDHSHLD	PCTSEVPROBNEED	PCTSEVMODPROBNEED	PCTOLDHSNEED	PCTMHNEED	PCTOWNINC	PCTRENTINS
Sample size (N)	118	118	118	118	118	118	118	118	118
Mean	0.016	0.055	0.324	0.005	0.020	0.052	0.005	0.954	0.243
Standard deviation	0.020	0.051	0.090	0.009	0.024	0.067	0.012	0.060	0.155
Coefficient of Variation	1.26	0.92	0.28	2.06	1.21	1.27	2.22	0.06	0.64
Maximum value	0.049	0.319	0.760	0.041	0.130	0.432	0.080	1.000	1.000
Minimum value	0.000	0.000	0.113	0.000	0.000	0.000	0.000	0.636	0.000
Range	0.049	0.319	0.647	0.041	0.130	0.432	0.080	0.364	1.000

variation and the range statistics indicate that, despite the low percentages, there is substantial variation in these conditions across metropolitan areas except for having homeowners' insurance.[14]

The author used the ordinary least squares (OLS) technique for the initial regression analysis; he eliminated some metropolitan areas from these regressions if it appeared that a zero-value for the dependent variable were the likely result of small samples.[15] The models reported in this paper were estimated using fractional logit. The author chose not to eliminate metropolitan areas from the fractional logit analysis because he considered that technique less affected by false-zeros.

Choice of Statistical Technique

Linking the AHS and the ACS should be thought of as a statistical, rather than an econometric, problem. The national AHS can produce a mean value for each of the seven AHS-based condition variables identified in Section II, and users could apply this mean value to all metropolitan areas in the United States as a first approximation. The statistical problems are (1) to find covariates to include in a model that would produce better results (smaller variance) than using the mean value for all metropolitan areas, and (2) to choose an appropriate technique to estimate the relationship between the dependent and independent variables. Section IV deals with the selection of covariates.

As noted in the discussion of small sample sizes, the author chose ordinary least squares for the initial analysis. The results were reported in the draft report of August 2009. While the F-statistics and the R^2s were favorable for eight of the nine regressions, the OLS technique has certain limitations that render it inappropriate for this purpose.[16] The two most serious limitations are:

- OLS equations can produce predictions of less than 0 percent or more than 100 percent for specific metropolitan areas. While the estimated OLS equations did not result in predictions outside of the 0-100 range for the 118 areas used to estimate the equations, there was no guarantee that inadmissible predictions would not result if the equations were used to make predictions for other metropolitan areas.

---

[14] The coefficient of variation is the ratio of the standard deviation to the mean, and the range is the difference between the maximum and minimum values.

[15] For each variable, the author estimated the ratio of the number of units in all 118 metropolitan areas with the specified characteristic to the number of occupied units in all metropolitan areas. The author then estimated the number of sample cases required to produce a probability equal to or less than 50 percent that no cases with the required characteristic would be obtained if the actual probability were equal to the computed ratio. If r is the ratio and n is the number of sample cases, the author found the minimum n that satisfied the following equation: $0.5 > (1-r)^n$. Then, for the model involving that dependent variable, the author eliminated all metropolitan areas where the observed percentage was 0 percent and the sample size was less than n. The rationale for using this procedure is that data from a metropolitan area should be used only if there is at least a 50-percent chance of finding the specified condition in the metropolitan area.

[16] The F-test statistic determines whether the set of independent variables contributes statistically to the prediction of the dependent variable, and the R^2 statistic measures the extent to which the regressions provide better predictions of the independent variables than the mean calculated over the areas included in the model.

- The large variation in sample sizes across the 118 metropolitan areas insured that errors in measuring a particular dependent variable would vary substantially across the metropolitan areas. This leads to a statistical problem called heteroskedasticity that produces inefficient, albeit unbiased, estimates.[17] An estimate is considered inefficient if it results in larger-than-necessary prediction errors.

At the suggestion of Dr. Lee, we chose to use fractional logit, a technique designed to estimate proportions.[18]

IV. Covariates for Modeling

The author included three types of covariates in models tested in Section V: ACS approximations of the AHS condition, covariates to distinguish among metropolitan areas on non-housing factors, and covariates related to local housing markets and economic conditions.

ACS Approximations of the AHS Condition

For most of the nine conditions described in Section II, approximations can be created using published ACS data. The starting point for each model is the ACS approximation. The bracketed numbers beginning with "B" are the ACS table numbers where information can be found.[19]

As the ACS approximation of PCTSEVPROB, the author defined:

> ACSPCTPROB = count of units without complete kitchen facilities [B25052] plus count of units without complete plumbing facilities [B25048])/count of occupied units [B25002]
>
> This variable created from the ACS does include units with many conditions that would be considered severe physical problems by the AHS. However, it will double count units that lack both complete kitchen facilities and complete plumbing facilities and will fail to count units with other physical deficiencies that the AHS considers severe.

The author will also use ACSPCTPROB as the best ACS approximation for PCTSEVMODPROB.

---

[17] Dr. Lee pointed out that the OLS results were also inefficient because they failed to take into account the correlation among the error terms in the nine regressions. A statistical technique called seemingly unrelated regressions (SUR) could eliminate this source of inefficiency but the underlying inefficiency resulting from heteroskedasticity would remain.

[18] "Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates" (with J.M. Wooldridge), *Journal of Applied Econometrics* 11, 619-632, November-December 1996.

[19] Those familiar with small area estimation will note that some of these ACS proxy variables are very similar to the small area estimates produced by "synthetic estimation." Synthetic estimation involves using known relationships among variables at a higher level of geography to estimate variables for small areas by applying those relationships to data available at the small level.

As the ACS approximation of PCTNEEDHSHLD, the author defined:

> ACSPCTNEED = (ACSNEED + ACSOTHNEED)/count of occupied housing units [B25002] where
>
> > ACSNEED = count of households with incomes below the poverty level plus count of households with incomes above the poverty level who are classified as "other families" plus the count of households with incomes above the poverty level who are married couple households with householders 65 years or older plus the count of households with incomes above the poverty level who are nonfamily households with householders of either sex 65 years or older [all in B17017]
> >
> > ACSOTHNEED = ((count of occupied units [B25002] – ACSNEED) * (entered in 2000 or later [from B05005]/total population [from B03001])

As the ACS approximation of PCTSEVPROBNEED, the author defined:

> ACSPCTPROBNEED = ACSPCTPROB* ACSPCTNEED

The author will also use ACSPCTPROBNEED as the best ACS approximation for PCTSEVMODPROBNEED.

As the ACS approximation of PCTOLDHSNEED, the author defined:

> ACSOLDHSNEED = ACSPCTNEED* (count of units built prior to 1940 [B25034]/total[B25034])

As the ACS approximation of PCTMHNEED, the author defined:

> ACSPCTMHNEED = ACSPCTNEED* (count of mobile homes[B25024]/total[B25024])

There are no ACS approximations for PCTOWNINS and PCTRENTINS. Table 4 presents the correlations between the AHS condition variables and these ACS approximations. The ACS approximation to physical problems does not correlate well with either severe or severe plus moderate as measured in the AHS. The variables derived from the AHS and ACS using age of housing have a correlation of 0.75.

Table 4: Correlations among AHS Condition Variables and their ACS Approximations

	PCTSEVPROB	PCTSEVMODPROB	PCTNEEDHSHLD	PCTSEVPROBNEED	PCTSEVMODPROBNEED	PCTOLDHSNEED	PCTMHNEED
ACSPCTPROB	0.0849	0.2284					
ACSPCTNEED			0.2749				
ACSPCTPROBNEED				-0.0156	0.2682		
ACSOLDHSNEED						0.7516	
ACSPCTMHNEED							0.3924

Covariates to Distinguish among Metropolitan Areas on Non-Housing Factors

Regional differences:

Three dichotomous variables with SOUTH as the exclude region:

NORTHEAST
MIDWEST
WEST

Size differences:

LNPOP = natural log of total population [B01003]

Racial and ethnic differences:

PCTBLACK = Black or African American alone (Hispanic plus nonHispanic)/total [B03002]

PCTHISP = Hispanic or Latino/total [B03002]

Age and growth history:

OLDMETRO = (count of units built prior to 1940 [B25034]/total [B25034])

FASTMETRO = (count of units built 1990 or later [B25034]/total [B25034])

Relative income:

RELINC = median household income [B19013]/$50,740 (U.S. median household income)

Education level of workforce:

EDUC = count of persons with associate's degree or higher (males + females)/total [B15002]

Covariates Related to Local Housing Markets and Economic Conditions

Vacancy rate:

VACRATE = vacant units/all housing units [B25002]

Homeownership rate:

HORATE = owner-occupied / total [B25003]

Cost of housing – 3 possibilities:

RELRENT = median gross rent [B25064]/$789 (U.S. median gross rent)

RELVALUE = median value [B25077]/$194,300 (U.S. median value)

OVERCROWD = count of units (owner + renter) with 1.01 or more occupants per room/total [B25014]

The author used only RELRENT and OVERCROWD in the models.

Type of housing structures:

UNITS2TO4 = (count of units in structures with 2 to 4 units/total [B25024])

UNITS50 = (count of units in structures with 50 or more units/total [B25024])

Appendix B provides the correlation matrix for the independent variables.

V. Statistical Models

This section reports regression models that use the covariates from Section IV to estimate the nine AHS-based condition measures identified in Section II. Our evaluation of the model focuses on two statistics:

- A Chi-squared test which determines whether the set of independent variables contributes statistically to the prediction of the dependent variables, and
- the R^2 statistic which measures the extent to which the regressions provide better predictions of the independent variables than the mean calculated over the areas included in the model.

In general, R^2 is not used to assess logit-type models; however, this statistic is appropriate in the context of this research project because the alternative to using a prediction from the fractional logit equation would be to use the mean of the dependent variable. After assessing these statistics, the author will examine the coefficients of the dependent variables primarily to determine whether there are any unwarranted relationships, but also out of curiosity.

Table 5 presents the Chi-squared and the R^2 statistics for the nine models.

Table 5: Chi-squared and R^2 Statistics

	Chi-squared	R^2
PCTSEVPROB	Not significant	0.118
PCTSEVMODPROB	Not significant	0.297
PCTNEEDHSHLD	Significant at 0.01	0.258
PCTSEVPROBNEED	Not significant	0.163
PCTSEVMODPROBNEED	Not significant	0.283
PCTOLDHSNEED	Not significant	0.685
PCTMHNEED	Not significant	0.533
PCTRENTINS	Significant at 0.01	0.339
PCTOWNINS	Not significant	0.298

In only two regressions is the relationship between the set of independent variables and the dependent variable considered statistically significant; these are the equations for the percentage of households in need (PCTNEEDHSLD) and the percentage of renters with insurance (PCTRENTINS). This is a very disappointing result. As noted earlier, eight of the nine OLS regressions were considered statistically significant. The added restriction that the predictions have to fall in the 0 to 100 range resulted in 6 of the 8 no longer being significant.

R^2 measures how much improvement the predictions of an equation make compared to the use of the sample mean. The R^2 statistics indicate that using the predictions from the fractional logit equations would eliminate 25.8 percent of the error resulting from using the mean estimate for PCTNEEDSHLD and 33.9 percent of the error from using the mean estimate for PCTRENTINS. In other words, the mean estimate, across the 118 metropolitan areas, of the percentage of households with one or more types of need (PCTNEEDHSHLD) is 32 percent (0.324). R^2 compares two alternatives: (1) the author could use the mean value of 32 percent for every metropolitan area, or (2) the author could use the value derived from the equation for PCTNEEDHSHLD. The R^2 for PCTNEEDHSHLD (0.258) means that the squared error from option (2) is 25.8 percent less than the squared error from option (1). The equation for PCTNEEDHSHLD furnishes a somewhat better estimate than just choosing the mean. These reductions in errors apply only to estimates for the 118 areas over which the equations were estimated.

While the equations for households in units built before 1940 who are in need (PCTOLDHSNEED) and for households in mobile homes who are in need (PCTMHNEED) are not statistically significant, the estimates from these equations would eliminate 68.5 percent and 53.3 percent, respectively, of the error resulting from using the mean estimates of these variables. The R^2s for five of the nine fractional logit equations are better than the R^2s from the OLS regressions even though the OLS technique maximizes the R^2, whereas the maximum likelihood techniques used in logit does not attempt to maximize R^2.

Table 6 contains the full regression models for the two statistically significant equations. (Appendix C contains the full regression models for the other seven AHS condition variables.) The author has shaded cells where the coefficient is statistically significant at the 10-percent level. There are only a few variables that are statistically significant even at the liberal 10-percent level. This is certainly the result of statistically significant relationships among the covariates. Appendix B displays a number of high one-on-one correlations, and the possibility exists that there are many high multiple correlations as well.

The lack of statistically significant parameter estimates is not a concern because, as noted at the beginning of Section IV, this is not an econometric analysis in which we are trying to uncover causal relationships among the variables. Our goal is more modest. We are looking for a set of covariates that *as a group* help predict the AHS condition measures. The Chi-squared and R^2 analyses in Table 5 are the essential tests for this objective.

The ACSPCTNEED variable was marginally significant in the equation for PCTNEEDHSHLD. PCTBLACK, UNITS2TO4, and the intercept were significant in this equation. LNPOP, OLDMETRO, VACRATE, and UNITS2TO4 were significant in the equation for PCTRENTINS. The PCTRENTINS equation did not have an equivalent ACS variable.

Table 7 contains predicted values for PCTNEEDHSHLD and PCTRENTINS for 27 metropolitan areas—11 areas from the sample used to estimate the equations and 16 areas not in the sample.[20] The predictions, as required by the logit specification, all fall in the 0 to 100 percent range. The predictions vary significantly from the mean value from the sample. For PCTNEEDHSHLD, the predictions range from 71.5 percent to 153.4 percent of the mean; for PCTRENTINS, the predictions range from 1.9 percent to 201.4 percent of the mean. In the sample, there was much greater variability in PCTNEEDHSHLD than in PCTRENTINS; the coefficients of variation were 0.28 and 0.64, respectively. Reliance on the equation rather than the mean would produce significantly different expectations for many of these metropolitan areas.

[20] The program used to estimate the fractional logit regressions (STATA©) does not provide confidence intervals for predictions for fractional logit.

Table 6: Regression Results

	PCTNEEDHSHLD				PCTRENTINS			
	Coefficient	**Standard Error**	**T-test**	**Probability > |t|**	**Coefficient**	**Standard Error**	**T-test**	**Probability > |t|**
ACSPCTNEED	3.0128	1.6971	1.78	0.076				
NORTHEAST	0.0639	0.1775	0.36	0.719	-0.2758	0.4042	-0.68	0.495
MIDWEST	-0.1148	0.1378	-0.83	0.405	-0.1560	0.2757	-0.57	0.572
WEST	-0.1764	0.1127	-1.57	0.117	0.0868	0.2610	0.33	0.74
LNPOP	0.0562	0.0348	1.61	0.107	0.1584	0.0817	1.94	0.053
PCTBLACK	-0.8717	0.4151	-2.1	0.036	-0.6482	0.9652	-0.67	0.502
PCTHISP	-0.4448	0.4112	-1.08	0.279	-0.8293	0.7454	-1.11	0.266
OLDMETRO	-0.7677	1.0343	-0.74	0.458	3.5568	1.8691	1.9	0.057
FASTMETRO	0.1031	0.6313	0.16	0.87	1.3440	1.4251	0.94	0.346
RELINC	0.2743	0.5741	0.48	0.633	0.8901	1.1945	0.75	0.456
EDUC	-0.6056	0.6755	-0.9	0.37	-3.3987	2.1668	-1.57	0.117
VACRATE	0.8298	1.7600	0.47	0.637	-9.1562	4.9741	-1.84	0.066
HORATE	0.5513	0.9017	0.61	0.541	-0.9178	2.3051	-0.4	0.691
UNITS2TO4	2.1679	1.1043	1.96	0.05	-8.4372	2.6428	-3.19	0.001
RELRENT	-0.2974	0.4094	-0.73	0.468	-0.7329	0.9511	-0.77	0.441
UNITS50	0.3274	3.0113	0.11	0.913	-7.8873	6.9356	-1.14	0.255
Intercept	-2.9493	1.2721	-2.32	0.02	-0.3938	2.0909	-0.19	0.851

Table 7: Predicted Values of PCTNEEDHSHLD and PCTRENTINS

Metropolitan Area	PCTNEEDHSHLD		PCTRENTINS	
In sample	**Prediction**	**Ratio to mean**	**Prediction**	**Ratio to mean**
Atlantic City, NJ	0.375	115.6%	0.098	40.5%
Beaumont-Port Arthur, TX	0.343	105.8%	0.253	104.0%
Charleston-North Charleston, SC	0.308	95.2%	0.170	69.8%
Corpus Christi, TX	0.339	104.7%	0.111	45.7%
Los Angeles-Long Beach-Santa Ana, CA	0.286	88.2%	0.205	84.2%
Memphis, TN-MS-AR	0.304	93.7%	0.225	92.5%
Miami-Fort Lauderdale-Pompano Beach, FL	0.384	118.4%	0.082	33.7%
New Orleans-Metairie-Kenner, LA	0.350	108.2%	0.097	39.7%
Oklahoma City, OK	0.351	108.3%	0.284	116.7%
San Francisco-Oakland-Fremont, CA	0.259	80.0%	0.226	92.8%
Washington-Arlington-Alexandria, DC-VA-MD-WV	0.258	79.5%	0.318	131.0%
Not in sample				
Anchorage, AK	0.282	87.1%	0.104	43.0%
Cedar Rapids, IA	0.273	84.3%	0.489	201.4%
Deltona-Daytona Beach-Ormond Beach, FL	0.400	123.3%	0.142	58.3%
Fairbanks, AK	0.232	71.5%	0.114	46.9%
Fargo, ND-MN	0.249	76.8%	0.341	140.2%
Gulfport-Biloxi, MS	0.358	110.5%	0.137	56.4%
Iowa City, IA	0.250	77.3%	0.263	108.1%
Key West, FL Micro Area	0.380	117.4%	0.005	1.9%
Naples-Marco Island, FL	0.497	153.4%	0.015	6.3%
Portland-Vancouver-Beaverton, OR-WA	0.272	83.9%	0.453	186.3%
Port St. Lucie, FL	0.412	127.2%	0.091	37.4%
Savannah, GA	0.299	92.4%	0.189	78.0%
Sioux Falls, SD	0.263	81.0%	0.392	161.1%
Springfield, MO	0.329	101.5%	0.384	158.2%
Virginia Beach-Norfolk-Newport News, VA-NC	0.290	89.5%	0.298	122.6%
Wilmington, NC	0.375	115.7%	0.095	39.1%
Sample Mean (unweighted)	0.324		0.243	
Minimum	0.232	71.5%	0.005	1.9%
Maximum	0.497	153.4%	0.489	201.4%

VI. Conclusions

As proposed, this research was exploratory. The results are only mildly encouraging. Of the nine equations estimated, only two were statistically meaningful using the Chi-squared test. However, the R^2s for these two equations indicate that using the predicted values may be a worthwhile improvement over assuming that all metropolitan areas have the same values, that is, using the sample mean. Out-of-sample predictions using these two equations suggest that conditions may vary substantially across metropolitan areas.

Equations with better statistical properties might result from using household-level data rather than metropolitan-level data. This would involve using microdata from the AHS to estimate the equations and microdata from the ACS public use files to use the equations for predictive purposes. Microdata analysis would require substantially more effort and would be technically more difficult. For example, using individual observations within metropolitan areas would have called for random effects modeling. Nevertheless, *if* the AHS is a unique source of information important in preparing for or responding to disasters, the effort and difficulty in small area estimation might be well worthwhile.

"If" is the key question. This research encountered an unexpected difficulty in identifying information unique to the AHS that would be obviously useful. The next major disaster may very well indicate what AHS information is crucial. Unfortunately, if such an event were to occur, there may not be time to carry out the desired microdata analysis. If so, the techniques explored in this paper might be a valuable second-best approach.

Appendix A: ACS and AHS Metropolitan Areas

	ACS METRO AREA (combined metropolitan areas in bold)	AHS occupied unit cases	Codes (SMSA) for corresponding AHS metropolitan areas
1	Akron, OH Metro Area	41	80
2	Albany-Schenectady-Troy, NY Metro Area	59	160
3	Albuquerque, NM Metro Area	64	200
4	Allentown-Bethlehem-Easton, PA-NJ Metro Area	57	240
5	Appleton, WI Metro Area	24	460
6	Atlanta-Sandy Springs-Marietta, GA Metro Area	268	520
7	Atlantic City, NJ Metro Area	9	560
8	Augusta-Richmond County, GA-SC Metro Area	28	600
9	Austin-Round Rock, TX Metro Area	90	640
10	Bakersfield, CA Metro Area	47	680
11	Baltimore-Towson, MD Metro Area	225	720
12	Baton Rouge, LA Metro Area	46	760
13	Beaumont-Port Arthur, TX Metro Area	15	840
14	Birmingham-Hoover, AL Metro Area	83	1100
15	Boston-Cambridge-Quincy, MA-NH Metro Area	362	1120, 4160, 7090
16	Boulder, CO Metro Area	28	1125
17	Bridgeport-Stamford-Norwalk, CT Metro Area	69	1160, 8040
18	Canton-Massillon, OH Metro Area	28	1320
19	Cape Coral-Fort Myers, FL Metro Area	11	2700
20	Charleston-North Charleston, SC Metro Area	39	1440
21	Chattanooga, TN-GA Metro Area	42	1560
22	Chicago-Naperville-Joliet, IL-IN-WI Metro Area	1046	1600, 2960,620,3965,9991
23	Cincinnati-Middletown, OH-KY-IN Metro Area	120	1640
24	Cleveland-Elyria-Mentor, OH Metro Area	204	1680
25	Colorado Springs, CO Metro Area	44	1720
26	Columbia, SC Metro Area	51	1760
27	Columbus, OH Metro Area	152	1840
28	Corpus Christi, TX Metro Area	38	1880
29	Dallas-Fort Worth-Arlington, TX Metro Area	467	1920, 2800
30	Davenport-Moline-Rock Island, IA-IL Metro Area	35	1960
31	Denver-Aurora, CO Metro Area	91	2080
32	Des Moines-West Des Moines, IA Metro Area	28	2120
33	Detroit-Warren-Livonia, MI Metro Area	585	2160
34	Duluth, MN-WI Metro Area	21	2240
35	El Paso, TX Metro Area	79	2320
36	Erie, PA Metro Area	13	2360
37	Eugene-Springfield, OR Metro Area	26	2400

	ACS METRO AREA (combined metropolitan areas in bold)	AHS occupied unit cases	Codes (SMSA) for corresponding AHS metropolitan areas
38	Evansville, IN-KY Metro Area	13	2440
39	Flint, MI Metro Area	36	2640
40	Fort Wayne, IN Metro Area	26	2760
41	Fresno, CA Metro Area	60	2840
42	Grand Rapids-Wyoming, MI Metro Area	68	3000
43	Greensboro-High Point, NC Metro Area	69	3120
44	**Greenville-Spartanburg-Anderson, SC CSA**	46	3160
45	Hartford-West Hartford-East Hartford, CT Metro Area	18	3280
46	Honolulu, HI Metro Area	75	3320
47	Houston-Sugar Land-Baytown, TX Metro Area	344	3360
48	Indianapolis-Carmel, IN Metro Area	120	3480
49	Jackson, MS Metro Area	22	3560
50	Jacksonville, FL Metro Area	102	3600
51	**Johnson City- Kingsport-Bristol (Tri-Cities) TN-VA Combined Metro Area**	20	3660
52	Kansas City, MO-KS Metro Area	174	3760
53	Knoxville, TN Metro Area	49	3840
54	Lakeland, FL Metro Area	14	3980
55	Lancaster, PA Metro Area	23	4000
56	Lansing-East Lansing, MI Metro Area	21	4040
57	Las Vegas-Paradise, NV Metro Area	122	4120
58	Lexington-Fayette, KY Metro Area	49	4280
59	Little Rock-North Little Rock-Conway, AR Metro Area	46	4400
60	Los Angeles-Long Beach-Santa Ana, CA Metro Area	1271	4480, 360
61	Madison, WI Metro Area	31	4720
62	McAllen-Edinburg-Mission, TX Metro Area	43	4880
63	Memphis, TN-MS-AR Metro Area	88	4920
64	Miami-Fort Lauderdale-Pompano Beach, FL Metro Area	534	5000, 2680, 8960
65	Milwaukee-Waukesha-West Allis, WI Metro Area	164	5080
66	Minneapolis-St. Paul-Bloomington, MN-WI Metro Area	277	5120
67	Mobile, AL Metro Area	30	5160
68	Modesto, CA Metro Area	29	5170
69	Montgomery, AL Metro Area	24	5240
70	Nashville-Davidson--Murfreesboro--Franklin, TN Metro Area	86	5360

	ACS METRO AREA (combined metropolitan areas in bold)	AHS occupied unit cases	Codes (SMSA) for corresponding AHS metropolitan areas
71	New Haven-Milford, CT Metro Area	54	5480
72	New Orleans-Metairie-Kenner, LA Metro Area	95	5560
73	New York-Northern New Jersey-Long Island, NY-NJ-PA Metro Area	2483	5600, 5380,875,3640,5015,5190, 5640,9992,9993
74	Oklahoma City, OK Metro Area	122	5880
75	Omaha-Council Bluffs, NE-IA Metro Area	56	5920
76	**Orlando-Deltona-Daytona Beach, FL CSA**	151	5960, 2020
77	Oxnard-Thousand Oaks-Ventura, CA Metro Area	68	6000
78	Palm Bay-Melbourne-Titusville, FL Metro Area	26	4900
79	Pensacola-Ferry Pass-Brent, FL Metro Area	17	6080
80	Peoria, IL Metro Area	32	6120
81	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Metro Area	631	6160
82	Phoenix-Mesa-Scottsdale, AZ Metro Area	331	6200
83	Pittsburgh, PA Metro Area	224	6280, 845
84	Providence-New Bedford-Fall River, RI-MA Metro Area	73	6480
85	**Raleigh-Durham-Cary, NC CSA**	85	6640
86	Riverside-San Bernardino-Ontario, CA Metro Area	190	6780
87	Rochester, NY Metro Area	85	6840
88	Rockford, IL Metro Area	15	6880
89	Sacramento--Arden-Arcade--Roseville, CA Metro Area	156	6920
90	Salinas, CA Metro Area	18	7120
91	Salt Lake City, UT Metro Area	130	7160
92	San Antonio, TX Metro Area	155	7240
93	San Diego-Carlsbad-San Marcos, CA Metro Area	309	7320
94	San Francisco-Oakland-Fremont, CA Metro Area	481	7360, 5775
95	San Jose-Sunnyvale-Santa Clara, CA Metro Area	171	7400
96	Santa Barbara-Santa Maria-Goleta, CA Metro Area	22	7480
97	Santa Rosa-Petaluma, CA Metro Area	31	7500
98	Sarasota-Bradenton-Venice, FL Metro Area	25	7510
99	Scranton--Wilkes-Barre, PA Metro Area	52	7560
100	Seattle-Tacoma-Bellevue, WA Metro Area	294	7600, 8200
101	Shreveport-Bossier City, LA Metro Area	28	7680
102	Spokane, WA Metro Area	29	7840
103	Springfield, MA Metro Area	70	8000
104	St. Louis, MO-IL Metro Area	226	7040, 2285, 275

	ACS METRO AREA (combined metropolitan areas in bold)	AHS occupied unit cases	Codes (SMSA) for corresponding AHS metropolitan areas
105	Stockton, CA Metro Area	43	8120
106	Syracuse, NY Metro Area	40	8160
107	Tampa-St. Petersburg-Clearwater, FL Metro Area	225	8280
108	Toledo, OH Metro Area	69	8400
109	Trenton-Ewing, NJ Metro Area	18	8480
110	Tucson, AZ Metro Area	77	8520
111	Tulsa, OK Metro Area	59	8560
112	Utica-Rome, NY Metro Area	12	8680
113	Vallejo-Fairfield, CA Metro Area	29	8720
114	Washington-Arlington-Alexandria, DC-VA-MD-WV Metro Area	406	8840
115	Waterbury, CT Metropolitan NECTA	14	8880
116	Wichita, KS Metro Area	49	9040
117	Worcester, MA Metro Area	31	9240
118	Youngstown-Warren-Boardman, OH-PA Metro Area	43	9320

Appendix B: Correlation Matrix for Independent Variables

	ACSPCTPROB	ACSPCTNEED	ACSPCTPROBNEED	ACSOLDHSNEED	ACSPCTMHNEED	NORTHEAST	MIDWEST	WEST	LNPOP	PCTBLACK	PCTHISP
ACSPCTPROB	1.0000										
ACSPCTNEED	0.0385	1.0000									
ACSPCTPROBNEED	0.9804	0.2148	1.0000								
ACSOLDHSNEED	-0.1826	-0.0728	-0.1980	1.0000							
ACSPCTMHNEED	0.2104	0.5509	0.3137	-0.4643	1.0000						
NORTHEAST	-0.0790	-0.1078	-0.0950	0.7105	-0.3030	1.0000					
MIDWEST	-0.2947	-0.0197	-0.2935	0.3327	-0.2856	-0.2595	1.0000				
WEST	0.1612	-0.4400	0.0743	-0.3523	-0.1413	-0.2535	-0.3038	1.0000			
LNPOP	-0.1601	-0.1814	-0.1929	-0.0875	-0.2560	-0.0306	-0.0775	0.0761	1.0000		
PCTBLACK	-0.0077	0.3613	0.0578	-0.2080	0.2880	-0.1404	-0.0965	-0.3995	0.0904	1.0000	
PCTHISP	0.3273	-0.1549	0.3156	-0.3932	0.0234	-0.1549	-0.3198	0.4417	0.1782	-0.3619	1.0000
OLDMETRO	-0.1769	-0.1363	-0.2029	0.9964	-0.4877	0.7184	0.3278	-0.3306	-0.0712	-0.2242	-0.3782
FASTMETRO	0.0179	0.2346	0.0698	-0.7618	0.4914	-0.5525	-0.2029	0.1280	0.0775	0.1347	0.2316
RELINC	-0.2025	-0.5451	-0.3092	0.0752	-0.5767	0.1860	-0.0877	0.3439	0.3915	-0.1714	0.0242
EDUC	-0.1011	-0.3114	-0.1846	0.1339	-0.4377	0.1285	-0.0084	0.1448	0.3364	-0.1097	-0.2119
VACRATE	0.0814	0.4432	0.1613	-0.2938	0.5374	-0.1545	-0.1103	-0.2718	-0.0906	0.2205	-0.0122
HORATE	-0.2542	0.2494	-0.1993	0.1456	0.2363	0.0266	0.3885	-0.5605	-0.2545	0.1371	-0.4752
UNITS2TO4	0.0060	-0.1630	-0.0205	0.7055	-0.3768	0.7401	-0.0666	-0.1387	-0.0010	-0.1872	-0.0509
UNITS50	0.0655	-0.2941	0.0027	0.0272	-0.4296	0.1304	-0.1063	0.1815	0.5038	-0.1343	0.2093
RELRENT	-0.0077	-0.4463	-0.0967	-0.2264	-0.3156	0.0544	-0.3737	0.5177	0.3771	-0.1282	0.3211
RELVALUE	0.0311	-0.5016	-0.0680	-0.1236	-0.3557	0.0228	-0.2956	0.6625	0.2914	-0.2584	0.3417
OVERCROWD	0.3269	-0.1470	0.3187	-0.3854	0.0329	-0.2072	-0.2968	0.5396	0.1637	-0.2338	0.8168

	OLDMETRO	FASTMETRO	RELINC	EDUC	VACRATE	HORATE	UNITS2TO4	UNITS50	RELRENT	RELVALUE	OVERCROWD
ACSPCTPROB											
ACSPCTNEED											
ACSPCTPROBNEED											
ACSOLDHSNEED											
ACSPCTMHNEED											
NORTHEAST											
MIDWEST											
WEST											
LNPOP											
PCTBLACK											
PCTHISP											
OLDMETRO	1.0000										
FASTMETRO	-0.7713	1.0000									
RELINC	0.1092	-0.1714	1.0000								
EDUC	0.1516	-0.0939	0.7353	1.0000							
VACRATE	-0.3142	0.3637	-0.3892	-0.3586	1.0000						
HORATE	0.1307	0.0715	-0.2564	-0.1435	0.3359	1.0000					
UNITS2TO4	0.7237	-0.5815	0.1755	0.1123	-0.2018	-0.1987	1.0000				
UNITS50	0.0513	-0.2270	0.4930	0.4233	-0.1455	-0.4079	0.1914	1.0000			
RELRENT	-0.1932	-0.0270	0.8131	0.4758	-0.0975	-0.4852	0.0918	0.5439	1.0000		
RELVALUE	-0.0894	-0.1790	0.7680	0.4330	-0.3003	-0.6102	0.1289	0.5100	0.9173	1.0000	
OVERCROWD	-0.3714	0.1293	0.1267	-0.1993	-0.0777	-0.6225	-0.0527	0.2601	0.4418	0.5322	1.0000

Appendix C: Fractional Logit Estimates for the Other Dependent Variables

	PCTSEVPROB				PCTSEVMODPROB			
	Coefficient	**Standard Error**	**T-test**	**Probability > |t|**	**Coefficient**	**Standard Error**	**T-test**	**Probability > |t|**
ACSPCTNEED								
ACSPCTPROB	-5.2664	35.6593	-0.15	0.883	10.4478	26.3435	0.4	0.692
NORTHEAST	0.4281	0.5417	0.79	0.429	-0.4338	0.3849	-1.13	0.26
MIDWEST	0.5677	0.4539	1.25	0.211	-0.3751	0.2980	-1.26	0.208
WEST	0.7082	0.3522	2.01	0.044	-0.2801	0.2614	-1.07	0.284
LNPOP	-0.0177	0.1241	-0.14	0.887	0.0334	0.0926	0.36	0.718
PCTBLACK	5.1903	1.4662	3.54	0	2.2785	0.7310	3.12	0.002
PCTHISP	2.5612	1.0983	2.33	0.02	0.4296	0.9169	0.47	0.639
OLDMETRO	2.8387	3.1496	0.9	0.367	0.2308	1.9978	0.12	0.908
FASTMETRO	1.4299	1.6027	0.89	0.372	-1.3200	1.4928	-0.88	0.377
RELINC	-2.2034	1.8627	-1.18	0.237	0.3813	1.5509	0.25	0.806
EDUC	3.5406	3.0584	1.16	0.247	-1.3407	2.7121	-0.49	0.621
VACRATE	-2.9247	5.8339	-0.5	0.616	-3.2564	3.6568	-0.89	0.373
HORATE	0.6308	3.0541	0.21	0.836	0.5736	2.2759	0.25	0.801
UNITS2TO4	0.0168	4.1944	0	0.997	1.4681	2.7292	0.54	0.591
RELRENT	0.6795	1.4864	0.46	0.648	-1.0477	1.0508	-1	0.319
UNITS50	1.2433	7.9552	0.16	0.876	10.9556	6.2848	1.74	0.081
Intercept	-5.9021	2.9175	-2.02	0.043	-2.6052	2.3264	-1.12	0.263

	PCTSEVPROBNEED				PCTSEVMODPROBNEED			
	Coefficient	**Standard Error**	**T-test**	**Probability > |t|**	**Coefficient**	**Standard Error**	**T-test**	**Probability > |t|**
ACSPCTPROBNEED	-65.8803	114.3013	-0.58	0.564	-7.9438	64.1697	-0.12	0.901
NORTHEAST	0.9001	0.9318	0.97	0.334	-0.7667	0.5078	-1.51	0.131
MIDWEST	1.0754	0.8667	1.24	0.215	-0.8555	0.4269	-2	0.045
WEST	-0.6787	0.8743	-0.78	0.438	-0.4610	0.3973	-1.16	0.246
LNPOP	-0.0265	0.1984	-0.13	0.894	0.0203	0.1214	0.17	0.867
PCTBLACK	4.0738	2.1168	1.92	0.054	1.7618	1.1591	1.52	0.129
PCTHISP	0.0839	1.9244	0.04	0.965	0.7265	1.2090	0.6	0.548
OLDMETRO	7.5726	4.0854	1.85	0.064	3.9475	2.8851	1.37	0.171
FASTMETRO	7.0776	2.3055	3.07	0.002	-1.0361	1.7367	-0.6	0.551
RELINC	-7.2507	2.9273	-2.48	0.013	-0.1288	2.3289	-0.06	0.956
EDUC	7.6935	5.2743	1.46	0.145	-0.6612	3.3636	-0.2	0.844
VACRATE	-8.3265	7.4277	-1.12	0.262	-0.3342	4.6626	-0.07	0.943
HORATE	6.1371	4.3357	1.42	0.157	0.8857	2.5708	0.34	0.73
UNITS2TO4	-0.2073	6.4485	-0.03	0.974	1.0894	3.9898	0.27	0.785
RELRENT	5.8123	2.2975	2.53	0.011	-0.9302	1.6668	-0.56	0.577
UNITS50	22.7393	13.6065	1.67	0.095	14.0666	7.0925	1.98	0.047
Intercept	-13.6801	4.4343	-3.09	0.002	-4.1594	2.6056	-1.6	0.11

	PCTOLDHSNEED				PCTMHNEED			
	Coefficient	**Standard Error**	**T-test**	**Probability > |t|**	**Coefficient**	**Standard Error**	**T-test**	**Probability > |t|**
ACSPCTOLDHSNEED	22.4424	20.8664	1.08	0.282				
ACSPCTMHNEED					41.8080	6.0764	6.88	0
NORTHEAST	1.0882	0.4738	2.3	0.022	0.6048	1.0390	0.58	0.561
MIDWEST	0.7693	0.3374	2.28	0.023	1.5797	0.8319	1.9	0.058
WEST	0.4765	0.3704	1.29	0.198	0.6787	0.5013	1.35	0.176
LNPOP	-0.1104	0.0940	-1.17	0.241	0.7471	0.1879	3.98	0
PCTBLACK	0.7858	1.3037	0.6	0.547	-0.3954	2.2538	-0.18	0.861
PCTHISP	-0.3845	1.1304	-0.34	0.734	-2.0096	2.2808	-0.88	0.378
OLDMETRO	-7.8653	9.7669	-0.81	0.421	-17.1197	8.7563	-1.96	0.051
FASTMETRO	-3.9265	1.7599	-2.23	0.026	-5.9456	1.9658	-3.02	0.002
RELINC	2.7880	1.5706	1.78	0.076	3.5614	2.3570	1.51	0.131
EDUC	-0.0555	2.2861	-0.02	0.981	-14.0630	3.8358	-3.67	0
VACRATE	6.5825	4.4638	1.47	0.14	-9.9068	10.4712	-0.95	0.344
HORATE	-2.0760	2.9284	-0.71	0.478	-5.6974	6.4108	-0.89	0.374
UNITS2TO4	-0.5221	2.3317	-0.22	0.823	5.4033	11.6386	0.46	0.642
RELRENT	-2.9447	1.2526	-2.35	0.019	0.8997	1.9428	0.46	0.643
UNITS50	-0.9985	10.3088	-0.1	0.923	-27.3179	16.6719	-1.64	0.101
Intercept	-0.6942	2.6467	-0.26	0.793	-8.1955	5.0650	-1.62	0.106

	PCTOWNINS			
	Coefficient	**Standard Error**	**T-test**	**Probability > |t|**
NORTHEAST	0.1614	0.6802	0.24	0.812
MIDWEST	0.9892	0.5594	1.77	0.077
WEST	0.7828	0.3769	2.08	0.038
LNPOP	-0.1438	0.1111	-1.29	0.195
PCTBLACK	0.6589	1.2659	0.52	0.603
PCTHISP	-0.4942	0.8924	-0.55	0.58
OLDMETRO	4.5449	4.0289	1.13	0.259
FASTMETRO	5.4160	2.8111	1.93	0.054
RELINC	-1.1439	2.0336	-0.56	0.574
EDUC	-0.4740	4.0203	-0.12	0.906
VACRATE	-11.2206	4.6812	-2.4	0.017
HORATE	1.2656	3.4849	0.36	0.716
UNITS2TO4	-1.3065	4.3789	-0.3	0.765
RELRENT	1.6425	1.5697	1.05	0.295
UNITS50	-7.0673	6.7807	-1.04	0.297
Intercept	3.0960	2.8048	1.1	0.27