

Data Shop

Data Shop, a department of Cityscape, presents short articles or notes on the uses of data in housing and urban research. Through this department, PD&R introduces readers to new and overlooked data sources and to improved techniques in using well-known data. The emphasis is on sources and methods that analysts can use in their own work. Researchers often run into knotty data problems involving data interpretation or manipulation that must be solved before a project can proceed, but they seldom get to focus in detail on the solutions to such problems. If you have an idea for an applied, data-centric note of no more than 3,000 words, please send a one-paragraph abstract to David.A.Vandenbroucke@hud.gov for consideration.

The CHAS Data: Obtaining Estimates of Housing Market Affordability

Angela M. Williams Foster
University of Pittsburgh

Abstract

The U.S. Department of Housing and Urban Development's (HUD's) Comprehensive Housing Affordability Strategy (CHAS) data provide vital information on housing affordability measures. This article presents potential uses of the CHAS data, as well as limitations of the data. Researchers should explore the benefits of using CHAS data to examine the role of affordable housing market structure on the persistence of urban issues.

Introduction

Obtaining housing affordability supply-and-demand estimates is important for assessing structural features of metropolitan housing markets. Unfortunately, decennial census aggregate-level data do not provide adequate measures of this vital housing market feature. The U.S. Department of Housing and Urban Development's (HUD's) Comprehensive Housing Affordability Strategy (CHAS) data

address this limitation.¹ The data provide special tabulations of census data for selected household and housing unit variables. The variables provided in the data enable users to characterize the affordable housing market for renter-occupied and owner-occupied housing units across local jurisdictions. The publicly available data, widely used by public and nonprofit agencies, are currently underused in the academic arena. This article provides an overview of the CHAS data and the potential uses for exploring housing and housing-related issues in metropolitan regions.

The first section, Data Overview, briefly describes the CHAS data. The second section, Data Uses and Limitations, addresses how to use the data to obtain measures of housing market characteristics. The third section, Conclusions, summarizes the benefits of using the CHAS data to develop neighborhood viability and metropolitan housing market measures.

Data Overview

The CHAS data were created to help jurisdictions meet requirements of the National Affordable Housing Act (HOME Program) of 1991. The Act requires eligible jurisdictions to submit a Consolidated Plan that presents a comprehensive housing affordability strategy for their geographic coverage area. The publicly available decennial census data were insufficient to provide summary measures of the status of the local affordable housing market. The special tabulations provided through the CHAS data enabled local governments to not only characterize the current status of the local affordable housing market but also to complete their Consolidated Plan. Compliance is necessary for jurisdictions to access federal funding.

The National Affordable Housing Act² furthers HUD's goal of ensuring that Americans have access to an adequate supply of decent, affordable housing. The Act is not limited to rental housing. Funding, incentives, and policy initiatives further the national housing policy goal of improving access to affordable homeownership opportunities as well. Although the CHAS data provide tabulations for households of varying income ranges, the policy goals focus on the inadequate and diminishing supply of affordable housing for low-income families, particularly those in the very low-income range.³ Federal funds to help jurisdictions provide safe, decent, affordable housing are available for participating jurisdictions (*Federal Register*, 2004).

Estimating Housing Unit Affordability

Analyses of affordable housing supply define affordable housing units as those for which households pay no more than 30 percent of their income on housing costs. The 30-percent threshold assures that households do not have an excessive rent burden and that they have income available

¹ For the 2000 data, see <http://www.huduser.org/datasets/cp.html> (accessed March 23, 2007). Data for 1990 are available through the State of the Cities Data Systems (SOCDS). SOCDS can be accessed through <http://socds.huduser.org/index.html> (accessed March 23, 2007).

² Title II (1990) 42 U.S.C. 12701.

³ Income classifications are as follows: extremely low-income (households earning less than or equal to 30 percent of HUD Area Median Family Income [HAMFI]), very low-income (households earning less than or equal to 50 percent of HAMFI), and low-income (households earning less than or equal to 80 percent of HAMFI).

for nonhousing living expenses. This affordability threshold is also a payment standard used by HUD and the U.S. Department of Agriculture. The CHAS data follow this threshold standard.

CHAS data provide tabulations at the block group,⁴ tract, county subdivision (that is, minor civil division), county, state, and national levels. The data use HUD-defined area median family income (HAMFI) classifications, selected housing-unit characteristics, and selected attributes of households residing in those units. Primary affordability variables are income range as a percentage of HAMFI and housing-unit affordability. The data also permit summary counts by household attributes such as race and Hispanic origin, poverty status, household type (elderly or family), presence of children, and college enrollment. The data also permit limited analyses of mobility or self-care limitations in elderly households. Housing-unit variables include tenure (renter-occupied or owner-occupied units), mortgage status, number of bedrooms (zero or one, two, or three or more), year structure was built, rent asked, and sales price asked. Tables also enable estimates of households by income range, occurrence of housing problems such as overcrowding (more than one person per room), housing cost burdens (paying in excess of 30 percent of income on housing costs), and limited housing quality measures (lack of complete plumbing or kitchen facilities).

A primary benefit of the CHAS data is that they contain special tabulations of census data categorized by household income (as a percentage of HAMFI) and housing affordability (as a percentage of HAMFI). These tabulations allow for consistent estimations of housing affordability and housing-cost burdens across several studies. Census Summary File (SF) income, rental, and housing value categories do not match with HUD-specified income limits and subsequent housing-affordability ranges. For example, consider a HUD income limit of \$37,500 for a low-income family of four. The maximum affordable rent for this family, \$892, falls within the census SF rental category of \$750 to \$999. In the absence of the special tabulations provided in the CHAS data, researchers would need to use potentially varying algorithms to estimate the number of households and housing units within desired affordability ranges. Although the Public Use Microdata Samples (PUMS) files alleviate this limitation, the files are limited in supporting analyses within metropolitan regions (that is, at lower levels of geography).

Data Uses and Limitations

Because of national housing policies, the CHAS data play a key role in the public and nonprofit sectors. The data also contain useful housing market measures that are relevant in the academic arena. Characterizing housing market affordability, particularly the relative mismatch between the supply of and demand for affordable housing units, is an important factor in studying metropolitan regions. Further, the extent to which households experience housing cost burdens is a key policy issue that warrants further research.

Selected Potential Analyses

The CHAS data include several key housing market variables that would support numerous analyses at varying geography levels. The multitude of potential analyses cannot be exhausted in

⁴ Given the low level of geography, block group tabulations present a limited number of variables.

a brief data note. The discussion that follows highlights a few key analyses based on two primary variables: household income range and housing unit affordability. As noted previously, the CHAS data policies focus on affordable housing opportunities for low-income households. The CHAS data provide specially tabulated summaries for five income classifications: (1) less than or equal to 30 percent HAMFI, (2) 30.1 through 50 percent HAMFI, (3) 50.1 through 80 percent HAMFI, (4) 80.1 through 95 percent HAMFI, and (5) greater than 95 percent HAMFI. Rent affordability is based on gross rent (that is, including utilities). The data provide estimates of the number of rental housing units affordable to households with incomes at or below the range of the specified HAMFI. Owner affordability tables provide affordable housing value ranges for costs associated with house purchases at the time the census was conducted.⁵

As with the census SFs, users have limited access to variable interactions. Fortunately, the data provide tables with multiple variable cross-tabulations. On the one hand, these multiple cross-tabulations often result in tables with many cells within a geographic unit (ranging between 8 and 96). On the other hand, the purpose of the data ensures that relevant variable interactions have been included. For example, tables A1A and A1B provide housing-unit summary counts for owner-occupied and renter-occupied housing units, respectively. The tabulations include variables for whether the unit has at least one housing-unit problem for four problem measures: (1) a lack of complete plumbing facilities, (2) a lack of complete kitchen facilities, (3) overcrowding (more than one person per room), or (4) cost burdened (paying more than 30 percent of income on housing costs). The tabulation further characterizes the race and Hispanic origin of the housing unit occupants (eight categories), along with low-income classification (five categories) separately for units with at least one housing problem or no housing problems. The resultant 80 cells for each geographic unit, while cumbersome, permits multiple estimates that would be unavailable through the census SFs. By aggregating across relevant variables, users can obtain the racial distribution of cost-burdened households, race-specific homeownership rates among cost-burdened households, or race or income distributions of households residing in units with housing problems. Aggregating the table also allows for estimates of the propensity of housing-unit problems by race or the distribution of housing-unit problems across income classifications within racial classifications.

A user can compute additional housing market viability and inequality measures by comparing across subcategories (for example, ratios, percentage differences, and gaps). Clearly this list is not exhaustive, but it illustrates the richness of the data in characterizing housing market viability. Research on race and income disparities in housing market outcomes would be further enhanced through analysis of CHAS data.

Tables A10B and A10C give the estimated number of households and housing units by affordability range, number of bedrooms, and income range for owner-occupied and renter-occupied units, respectively. Owner-occupied housing tables differ slightly from renter-occupied housing tables in the lowest affordability category reported. Although renter-occupied housing tabulations provide the number of rental units affordable at the extremely low-income threshold, the lowest reported affordability range for owner-occupied tabulations is the very low-income threshold. Given the

⁵ The affordability range assumes an interest rate of 7.9 percent and uses national averages for utilities, taxes, and homeowner's insurance. Owner affordability tables use a multiplier, 2.9, to determine the house value affordability threshold.

expected relatively low ownership propensity among extremely low-income households, this difference is not a significant limitation of the data. Including measures of unit size provides a more comprehensive depiction of the affordable housing market. For instance, the data permit estimates of housing problems not directly captured in the cross-tabulations, such as the potential lack of affordable housing supply for large families.

Estimating the supply of affordable housing units by tenure and the number of households at varying income levels is not the only use of the CHAS data. The data provide housing market features that, when appended to individual-level data with geographic identifiers, would enrich current research performed on these data. Future research should use the CHAS data to examine the effect of affordable housing market features on household-level tenure and household mobility outcomes. Restricted access versions of the American Housing Survey (AHS) and Panel Study of Income Dynamics contain geographic identifiers that would support hierarchical models explicitly acknowledging that these household outcomes occur within the context of place.

Limitations

This section highlights three primary limitations of the CHAS data: (1) aggregate tables, (2) estimate errors from rounding rules, and (3) inadequate measures of housing quality. The latter two limitations are best addressed through obtaining estimates at the highest level of aggregation for both geography and table components. The discussion includes ways to address housing quality limitations by incorporating estimates from an external data source. The rounding scheme, although not mitigated, should be considered when interpreting estimated housing-unit and population counts. As a consequence of the limitations, users should exercise caution when making comparisons between CHAS data and census SF data. In addition, the data do not give estimates for metropolitan statistical areas (MSAs). Users can use census tract or county-mapping files to aggregate the CHAS-provided geography to the MSA level. Finally, combining the CHAS data with the GeoLytics® Neighborhood Change Database allows for analysis of housing market affordability for within the context of changing housing market conditions.

Aggregate Tables. As noted, the summary tables present limited interactions. For instance, the user cannot determine which housing-unit problem is present from tables A1A and A1B; however, this limitation is one that is also present in the census SFs. Users of decennial census data acknowledge that individual data, such as the PUMS, provide the most flexibility in characterizing household and housing-unit conditions. Despite this limitation, the policies motivating the CHAS data make many useful variable interactions available. As such, variables contained in the data characterize numerous housing market viability measures. The user can analyze external data sources (for example, PUMS) to determine the extent to which omitted interactions are correlated with obtainable interactions through CHAS.

Rounding Rules. CHAS tabulations use a rounding scheme for most tables. The rounding scheme results in internal file discrepancies and in differences between CHAS and publicly available decennial census count data.⁶ The total column for a particular table provides the total housing-unit

⁶ According to CHAS documentation, the rounding scheme has three rules: 0 counts remain 0, values 1 through 7 are reported as 4 (the midpoint), and remaining values round to the nearest multiple of 5.

count without applying the rounding scheme. Thus, internal discrepancies result from differences between the summed values of the rounded cell counts (computed total) and the nonrounded total (provided total). Comparing these two totals is beneficial and demonstrates the effect of rounding. Because of rounding, the computed totals also will differ from population and housing-unit counts available through the SFs. A user should make similar comparisons to estimate the rounding effect.

One way of addressing the rounding limitation is to provide bounded estimates of housing units or population counts for specified variables (for example, rental housing units affordable to low-income households). For small geographic levels with small housing-unit counts (for example, tracts or block groups) the bounded-estimate method should be used with caution. A user is strongly encouraged to compare these estimates to the census SFs and weigh the costs of disaggregating to a low level of geography. Another way of addressing the limitation is to use the data to obtain distributions across categories. The example at the end of this section demonstrates that the percentage distribution of affordable housing units or low-income households generally offsets the overestimations and underestimations of the rounding scheme. A user could then apply the estimated distributions to either the provided CHAS table totals or the census SF totals.

Housing Quality. A third limitation of the CHAS data is the lack of adequate measures for housing quality. The two-fold definition for affordable housing requires that housing costs are within the household's ability to pay, while leaving sufficient financial resources for other pertinent living expenses, *and* that the unit is decent and safe. CHAS data can be used to determine the number of housing units at various affordability levels that have at least one housing problem. The data are limited in identifying housing quality separate from crowding and cost burdens. Generally, affordable housing studies assess decency or safety standards using housing adequacy measures provided in the AHS. The AHS combines several variables capturing physical problems to create an index of housing adequacy: adequate, moderately inadequate, or seriously inadequate. These variables include major plumbing, heating, or upkeep deficiencies, as well as the presence of peeling paint, leaks, or rats. Because the CHAS data are based on decennial census data, the tabulations do not include these important quality measures; and thus, CHAS indicators are not sufficient to estimate the number of housing units with moderate or severe deficiencies.

One way to address this limitation for higher geography levels (for example, MSAs) is to supplement the CHAS data with AHS estimates. The AHS is limited in addressing housing quality at lower geography levels (for example, counties or census tracts). The national AHS does not permit analysis at the county level. Metropolitan files, which could permit county-level analyses, often are not conducted within desired timeframes. The national AHS does allow metropolitan regional estimates of housing quality differentials by urban or rural status, household income range, and housing costs. The AHS estimates approximate the distribution of moderately or severely inadequate housing units by housing cost range. To account for housing quality, adjust supply estimates from the CHAS data using AHS estimates of housing adequacy for varying housing-cost affordability levels.

Example

The following example explores the use of CHAS data to characterize Pennsylvania's affordable housing market. Exhibit 1 provides an example for table A10C at the state level (Summary Level 040r)

and tract level (Summary Level 080r).⁷ The table gives the number of renter-occupied housing units by household income range (five categories) and number of bedrooms (three categories: zero to one bedroom, two bedrooms, and three or more bedrooms) for four rental affordability levels (affordable for households in four income categories: less than or equal to 30 percent of HAMFI, 30.1 percent through 50 percent of HAMFI, 50.1 percent through 80 percent of HAMFI, or greater than 80 percent of HAMFI). The resulting table has $5 \times 3 \times 4 = 60$ cells for each jurisdiction.

The rounding scheme applied at the state level overestimates the number of housing units by five units. Disaggregating to the census tract level exacerbates the rounding errors with an underestimate of approximately 3,600 households or units. Although the differential in overall household and housing-unit counts is substantial, the distributions of housing units across affordability ranges and households across income ranges are comparable at the state and aggregated census tract levels.

In addition to demonstrating the rounding effect, exhibit 1 illustrates several useful affordability measures obtained through the CHAS data. Aggregating the data across the number of bedrooms permits analysis of household distributions across three housing-unit affordability outcomes: affordable housing occupancy, rent-burdened households, and households renting down. Affordable housing occupancy refers to households paying no more than 30 percent of income on housing costs. Rent-burdened households occupy housing units with costs in excess of the affordability threshold. Households renting down occupy housing units affordable at a lower affordability range. The last classification is pertinent in characterizing housing market affordability because those households further restrict the low-income affordable housing supply.

Conclusions

The previous discussion highlights ways to obtain useful distributions of housing market viability, such as housing cost burdens, the gap between affordable housing supply and demand, and the distribution of income levels within affordability ranges (for example, renting down). Current research on housing and urban issues would benefit from including the affordability measures available through this rich data. Future research should use CHAS data to examine the role of affordable housing market structure on the persistence of urban issues (for example, concentrated poverty, housing mobility differentials, and racial residential segregation).

The origins of CHAS data in the policy field should not hinder its use in academic scholarship. The data enable research contributions beyond assisting jurisdictions to prepare Consolidated Plans. The CHAS data provide an opportunity for bridging the gap between practitioner and academic research. Future research using the CHAS data should explore these linkages and ways to better inform and evaluate housing policy.

⁷ Summary levels 040r and 080r indicate that the data cells provide rounded estimates.

Exhibit 1

Example of CHAS Housing Affordability Analysis and the Effect of the Rounding Scheme (1 of 2)

		State Summary Level				Tract Summary Level			
		Affordability Threshold				Affordability Threshold			
Income Range	Less than or equal to 30 percent	30.1 through 50 percent	50.1 through 80 percent	Greater than 80 percent	Income Range	Less than or equal to 30 percent	30.1 through 50 percent	50.1 through 80 percent	Greater than 80 percent
≤ 30% HAMFI					≤ 30% HAMFI				
0 or 1 BR	62,820	53,190	34,965	7,715	0 or 1 BR	62,844	53,194	34,640	7,598
2 BR	33,735	40,375	22,605	3,075	2 BR	33,866	40,316	22,389	3,007
3 or more BR	34,795	27,480	12,235	1,635	3 or more BR	34,821	27,593	12,063	1,590
30.1 through 50% HAMFI					30.1 through 50% HAMFI				
0 or 1 BR	20,885	44,690	27,950	5,945	0 or 1 BR	20,926	44,653	27,810	5,901
2 BR	16,875	38,755	24,720	2,840	2 BR	16,984	38,720	24,548	2,793
3 or more BR	20,160	24,005	14,245	1,465	3 or more BR	20,083	24,093	14,123	1,406
50.1 through 80% HAMFI					50.1 through 80% HAMFI				
0 or 1 BR	9,260	43,825	42,855	7,935	0 or 1 BR	9,134	43,727	42,738	7,804
2 BR	14,265	51,400	44,415	4,605	2 BR	14,370	51,362	44,390	4,536
3 or more BR	21,005	32,010	24,450	2,545	3 or more BR	20,978	32,056	24,428	2,442
80.1 through 95% HAMFI					80.1 through 95% HAMFI				
0 or 1 BR	2,045	12,740	17,865	3,450	0 or 1 BR	1,964	12,465	17,656	3,351
2 BR	4,610	18,125	20,615	2,465	2 BR	4,738	18,052	20,486	2,343
3 or more BR	6,770	11,835	10,825	1,530	3 or more BR	6,974	11,841	10,613	1,451

Exhibit 1

Example of CHAS Housing Affordability Analysis and the Effect of the Rounding Scheme (2 of 2)

		State Summary Level				Tract Summary Level			
		Affordability Threshold				Affordability Threshold			
Income Range	Less than or equal to 30 percent	30.1 through 50 percent	50.1 through 80 percent	Greater than 80 percent	Income Range	Less than or equal to 30 percent	30.1 through 50 percent	50.1 through 80 percent	Greater than 80 percent
> 95% HAMFI					> 95% HAMFI				
0 or 1 BR	5,385	27,090	52,230	16,805	0 or 1 BR	5,360	27,040	52,214	16,684
2 BR	12,150	47,585	84,650	19,820	2 BR	12,217	47,507	84,640	19,561
3 or more BR	23,165	35,990	45,180	12,185	3 or more BR	22,931	35,879	45,113	12,095
Total	287,925	509,095	479,805	94,015	Total	288,190	508,498	477,851	92,562
Percent	21	37	35	7	Percent	21	37	35	7
	Computed Total	1,370,840			Computed Total	1,367,101			
	Provided Total	1,370,835			Provided Total	1,370,711			
	Difference	5			Difference	-3,610			

BR = bedroom.

CHAS = Comprehensive Housing Affordability Strategy.

HAMFI = HUD area median family income.

Author

Angela M. Williams Foster is professor of public and urban affairs at the Graduate School of Public and International Affairs, University of Pittsburgh.

Reference

Federal Register. 2004. "Department of Housing and Urban Development. 24 CFR Part 92, HOME Investment Partnerships Program; Amendments to Homeownership Affordability Requirements; Interim Rule." November 22. Vol. 69, no. 224: 68050–68052.