

Known Facts or Reasonable Assumptions? An Examination of Alternative Sources of Housing Data

David Listokin, Elvin Wyly, Ioan Voicu, and Brian Schmitt*

Abstract

Researchers can choose from a wide variety of data sources to measure the dynamics of production, consumption, and change in America's housing stock. Yet there have been few systematic comparisons of housing and household characteristics across data sets. Such considerations typically appear when differences in survey sampling, database content, or question wording are invoked to explain unexpected or inconsistent results.

We describe key differences of nine data sets commonly used in housing research and provide an empirical analysis of differences among the American Housing Survey, the Consumer Expenditure Survey, the National Survey of Families and Households, and the Survey of Income and Program Participation. The analysis compares selected financial and demographic characteristics by household tenure. Although demographic characteristics such as age of householder and household size generally are comparable across data sets, there are considerable and statistically significant differences for important financial characteristics.

Keywords: Data sources; Household surveys; Social measurement

Introduction

Scholars and policy makers who read *Journal of Housing Research* have an avid interest in the careful, rigorous empirical measurement of household and housing characteristics; thus it is common for articles in this journal to include a simple, straightforward description of such issues in a section titled "Data." Yet empirical measurement is neither simple nor straightforward. Complexity appears as soon as we try to define our terms:

data *pl* of DATUM.

datum n. a known fact || the assumption which forms the basis for an inference or a conclusion || a starting point from which, e.g., a survey is made. (Lexicon Publications 1990, 244–45).

* David Listokin is Professor at the Center for Urban Policy Research (CUPR) at Rutgers University. Elvin Wyly is Assistant Professor in the Department of Geography at the University of British Columbia, and Ioan Voicu and Brian Schmitt were Research Associates at CUPR.

The authors are most grateful to Patrick A. Simmons for his invaluable guidance, advice, and editing on this project; Kate Fullerton (Rutgers University) for providing excellent research assistance; and Shannon Darroch (Rutgers University) for typing and retyping many versions of the manuscript and tables. The authors assume all responsibility for any shortcomings.

The word “datum” is Latin for “given.”

In this article, we wish to explore the implications of this etymology for housing research. Econometric analysis and public policy are based on the conviction that objective, well-conceived, and carefully implemented surveys provide known facts that can help us understand the true state of the housing market. Yet an examination of alternative sources of housing information suggests that our data may only provide a starting point or a set of reasonable assumptions from which we may draw inferences.

The next section of this article provides background for this analysis and is followed by an overview of the range of data sets used in housing research. Then, an empirical analysis of several housing data sets compares values for a number of key variables. The article concludes with some thoughts on the implications of our findings for research and policy.

Background

Housing research is a rich and interdisciplinary field of inquiry with deep roots in scholarly work and public policy concerned with the economic facets of housing markets, the interdependencies of housing and sociodemographic trends, and the intricate geographical processes of metropolitan evolution and neighborhood change. For many years, empirical work in housing studies has drawn heavily from information sources that present a broad view of the conditions of population and housing. The most familiar sources include the summary files (SFs) and public use microdata samples (PUMS) from the decennial census, the Current Population Survey (CPS), and the American Housing Survey (AHS). These core data sets have anchored a broad body of research for many years, and they continue to provide the empirical base for innovative theoretical and policy-oriented studies (e.g., Boehm 1995; DiPasquale and Somerville 1995; Duda and Belsky 2000; Early 1998; Early and Phelps 1999; Galster et al. 1996; Kiel and Zabel 1997; Moore 1991; Nelson 1994; Pitkin 1998; Quercia et al. 1998; Rasmussen, Megbolugbe, and Morgan 1995; Wieand 1983).

In recent years, however, analysts have turned to an expanding array of data sources. Many scholars grow impatient waiting for the long-form census SFs and PUMS (which often are regarded as historical documents by the time they are released),¹ and traditional sources include little information on topics that have attracted great attention—household wealth, debt, and the relation between housing consumption and eligibility for various types of public assistance. Researchers also increasingly are interested in panel surveys specifically designed to illuminate change; with the exception of AHS, the core sources are cross-sectional,

¹ The U.S. Bureau of the Census recently launched the American Community Survey (ACS) to provide more timely data for researchers as well as state and local policy makers. ACS is based on a continuous measurement design and is intended to replace the long form of the decennial census in 2010. For areas with populations less than 20,000, ACS will provide five-year averages for sample characteristics. Areas with populations between 20,000 and 65,000 will be reported in three-year averages. Annual breakdowns will be available for areas with populations of more than 65,000. See the World Wide Web site <http://www.census.gov/acs/www>.

and even AHS provides no panel information on household change.² Through all of these changes, scholars working from the varied theoretical and empirical traditions of separate disciplines (economics, sociology, geography, urban planning, and public policy) have come to the table with distinctive sets of thematic needs, analytical priorities, and stylistic or methodological preferences.

A review of contemporary housing research points to the wide variety of information resources being used. Stegman et al. (1991) looked at the Panel Study of Income Dynamics (PSID) to consider the long-term effects of government housing subsidies. Boehm and Schlottmann (2000) used PSID to examine housing and wealth accumulation. Rosenthal (2000) used the Survey of Consumer Finances (SCF) to estimate the homeownership rates that would prevail if borrowing constraints were eliminated (see also Duca and Rosenthal 1994). Lang, Hughes, and Danielsen (1997) used a proprietary data source, Claritas, to consider target groups for marketing central city housing. Savage (1997, 1999) and Savage and Fronczek (1993), in the Who Can Afford To Buy a House series, assembled income, asset, and debt data from the Survey of Income and Program Participation (SIPP). Calhoun and Stark (1997) tapped the National Survey of Families and Households (NSFH) to examine the ability of renters to secure homeownership. Listokin et al. (2002) considered how underwriting data pertaining to income, assets, debt, and credit could be obtained from SIPP, NSFH, the Consumer Expenditure Survey (CES), and other sources. In recent years, there also seems to be an expanded effort to use multiple information sources to shed light on complicated empirical and theoretical questions. Studies by Green and White (1997) and Green (2000) used PUMS, PSID, and other data to analyze the effects of housing tenure choice on child outcomes. Pitkin (1998) tapped CPS, AHS, and the Housing Vacancy Survey to consider changes in homeownership and households.

The storehouse of data sets available to contemporary housing researchers is broader and richer than it was a generation ago. When undertaking empirical work, analysts are able (and often are required) to sift through a variety of data sets to make choices on survey topical coverage, geographic resolution and time period, and organization of survey results. A simple review of these issues should help researchers understand key differences and choose sources appropriate for their work.

Moreover, the existence of a wide variety of data sets with housing information raises an important question: Does the choice of data have any influence on our view of household circumstances and housing conditions? This issue has been discussed in several quarters of the literature. To cite only a few examples, Case, Pollakowski, and Wachter (1991) and Pollakowski (1995) considered the relative merits of house price information obtained from real estate agents, property assessment files, and other sources. Eller and Fraser (1995) calculated the net worth of households from SCF and SIPP and found striking differences: The 1993 mean net worth estimated from SIPP was \$99,772, whereas the 1992 mean net worth estimated with SCF (inflated to 1993 dollars) was \$226,900. Juster and Kuester (1991) compared the data on wealth from the SCF and the Retirement History Survey and found considerable

² The same housing units are surveyed in AHS every two years, but the households living in those housing units may change. AHS, therefore, is ideal for analyzing the flow of households living in a given housing unit, rather than for analyzing a given household over time.

differences. Unfortunately, systematic and specific comparisons such as these are rare; relatively little is known about the contrasting profiles of household and housing characteristics offered by various household surveys. This issue is far more than an obscure methodological detail. If the basic socioeconomic characteristics of households differ across separate data sources, it is likely that such variations will be altered and magnified by the widespread use of policy-oriented benchmarks and multivariate econometric models of housing market processes. This article provides a first step toward understanding the range of information now available to housing researchers, and it analyzes similarities and differences across these sources.

Objectives

Our first objective is to provide an overview of data sets that can be used to conduct housing research. The housing research field is broad, and the wide variety of sources that have been used range from the High School and Beyond Survey (Green and White 1997) to the National Longitudinal Survey of Youth (Haurin, Parcel, and Haurin 2002). Providing an encyclopedic compilation of every possible source that has housing research utility would be an impossible task. Instead, this overview focuses on a much more limited pool of publicly available data sets that increasingly are being used by contemporary housing researchers. We provide a descriptive review of the coverage and relative strengths of nine data sources that commonly are used in housing research or that include relatively detailed information on housing-related topics.³

Our second objective is to offer a preliminary empirical analysis of a subset of these data sets to determine whether alternative sources yield systematically different views of households and housing conditions. We devised a simple, illustrative analysis of the socioeconomic profile of renters and homeowners—a dichotomy that is a central concern in much work in the housing literature (e.g., Calhoun and Stark 1997; Listokin et al. 2002; Savage 1997, 1999; Savage and Fronczek 1993). We selected a comparable sample of households from four separate data sets, and we tabulated selected social-demographic, housing, and household financial characteristics. We examined variations in household characteristics by tenure, and we analyzed differences in key indicators across data sets. We sought to answer such questions as: Does the choice of data set have any substantial effect on estimated income or wealth or on contrasts between renters and owners?

³ It is important to recognize that our selection of data sets is far from comprehensive and emphasizes broad, publicly available sources that measure the demand side of the market. Our analysis does not include several alternative sources that can be used to measure the supply side or to examine specific housing submarkets. Examples include the Survey of Construction (available from the U.S. Bureau of the Census), the annual releases from the Home Mortgage Disclosure Act (from the Federal Financial Institutions Examination Council), several data sets related to Federal Housing Administration lending (the F42 EDS Case History File and the F42 BIA Composite File, both from the U.S. Department of Housing and Urban Development [HUD]), and the GSE Public Use Database, describing the loan purchases of Fannie Mae and Freddie Mac (and maintained by HUD). Private-sector data initiatives have proliferated in recent years, and many of these efforts involve the synthesis or enhancement of public data from the census and other sources. Other private initiatives offer specialized information not available elsewhere, such as the Existing Home Sales Survey (from the National Association of Realtors) and the Primary Mortgage Market Survey (from Freddie Mac).

Overview of the Data Sets

The data sets included in our review all are sponsored or conducted by public institutions. Most can be accessed at no cost via the World Wide Web, with a wide variety of additional data products available at nominal cost. Four of the data sets are broad surveys of housing and demographic topics: SFs and PUMS from the decennial census, the CPS Annual Demographic Survey, and AHS. The remaining five data sets provide more specialized coverage of subjects such as consumer expenditures or household finances: SIPP, CES, PSID, SCF, and NSFH. For ease of reference and comparison, all nine databases are described in tables in appendix A. Each table is divided into two sections. A profile section presents a snapshot of background information on the purpose of each survey, its sample characteristics, the time frame of the data collection, and the geographic scales at which each data set can be analyzed. A content section summarizes data characteristics for social and demographic information (household size and composition, age of householder, and race and ethnicity), household finances (income, assets, liabilities, employment history, credit history, mortgage information, and property taxes), and housing information (property and neighborhood characteristics, tenure status, and house value).

Profile Differences

It should come as no surprise that different data sources reflect competing priorities and concerns. The considerable expense of each question item on such large-scale surveys necessitates different kinds of justifications by research communities, statistical experts, and sponsoring agencies. AHS is particularly well-suited to the study of the changing conditions of the housing stock, CES sheds light on detailed aspects of consumption behavior, and SIPP provides unparalleled information on eligibility for various means-tested government programs. Yet a variety of data sets include a subset of shared indicators, and the possible uses of some surveys often are not apparent at first glance. NSFH, for instance, is used most commonly in demographic and sociological studies of household relationships, but Calhoun and Stark's (1997) innovative analysis illustrates its utility for housing research as well.

Contrasts in sample characteristics, periodicity, and geographic resolution often are interrelated and shape the value of different information sources for various types of inquiry. AHS, for instance, is released more frequently than the decennial census—but provides a much smaller sample size and a relatively coarse spatial resolution.⁴ Separate products from the decennial census illustrate similar kinds of trade-offs and compromises. PUMS provides detailed data on individual people, households, and families—units of analysis that are not available in the census SF tabulations. However, the geographic resolution of PUMS (national, state, and public use microdata area) is not nearly as finely grained as that of the census SFs.

Although the appendix A tables summarize these sorts of considerations, a host of other methodological and technical issues must be scrutinized for each data set. Many of the

⁴ SF tabulations provide coverage for certain variables down to the census tract or block level, whereas the metropolitan files of AHS offer tabulations and microdata only for a small number of subareas within metropolitan areas.

microdata samples, for instance, include codes for multiple and often overlapping units of analysis—greatly expanding analytical possibilities but creating innumerable possibilities for error when extracting data. There are often variables that describe the household and a separate set of variables that describe people in the household. For example, SIPP data can be analyzed by household, family, subfamily, or person. Analyzing easily understood sociological categories (e.g., white renters) requires extracting household-level information (tenure status) as well as reference person data (race/ethnicity).⁵ However, the researcher must make sure to use a consistent unit of analysis. If household information is to be attributed to everyone in the household, the researcher needs to select only one person from that household. Failure to properly identify the unit of analysis can easily create enormous errors. For example, a married couple with four children and \$100,000 a year in earnings can erroneously become six households, each with \$100,000 a year in earnings.

Content Differences

The second portion of tables A.1, A.2, and A.3 summarizes selected demographic, financial, and housing variables in the nine data sets. To illustrate key differences, we sketch out a brief review of four indicators—income, debt, assets, and house value.

Although all the sources provide some information on household income, they differ in the nature and detail of the income data. Both NSFH and SCF report income, but only at the ordinal scale. CES data are superior in that CES offers interval-scale data on income, as well as income by detailed source (nine categories).⁶ SIPP, however, offers the most income detail; it gives interval figures for 38 separate sources of income and reports income by person within the households (table A.2). Furthermore, SIPP’s sample size (20,000 households) is four times that of CES’s sample size (5,000 households.) Thus, because of its robust sample size and wealth of detail, SIPP offers numerous advantages to housing researchers seeking income data.

Debt provides a valuable measure of household financial constraints and opportunities and of past expenditure commitments. Several data sources (PUMS, CPS, and AHS) have little debt information of any kind. PUMS and AHS, for example, contain information on mortgage payments and/or outstanding mortgage debt but on no other debt data. Other sources include information with varied levels of specificity. NSFH has only ordinal-scale debt data. SIPP offers interval-scale data on outstanding debt owed in several categories: store bills and credit card debt, bank loans, auto loans, and other categories. Unfortunately, SIPP reports only total outstanding debt—not the periodic (monthly or weekly) debt repayment, which is of more

⁵ These different types of data usually are coded in separate files and are linked through a common field. Some data sets (e.g., SIPP) also offer a “rectangular” version of the data. These simple files are similar to the rows and columns of a spreadsheet and typically are much easier to work with.

⁶ Income source is of interest in housing finance because, over time, underwriters have accorded different significance to various forms of income, such as alimony, child support, and welfare. Historically, mortgages discounted income from alimony, child support, and welfare. Contemporary mortgage products are much more inclusive of those income resources.

interest in certain types of housing research.⁷ CES offers interval-scale data on periodic (monthly) debt expenditures and reports these costs by type: credit card debt, installment debt, medical bills, and others. CES also asks questions regarding “total amount owed this month” compared with “total amount owed last year.” Responses to these questions permit an estimate of annual changes in debt burdens—useful trend line information. For instance, mortgage underwriters consider whether a household’s debt has been increasing or decreasing. Thus, housing researchers interested in debt data might very well be attracted to CES because of the richness of the debt information it provides. Yet SIPP has a larger sample size than CES, and CES documentation suggests that debt information in this survey may not be as accurate as expenditure data.

Assets also are crucial in the measurement of household circumstances. Two of the data sets (the census SFs and PUMS) have no information on assets other than owner-occupied homes. The other data sets are more informative, but to different degrees. For example, SCF and NSFH report assets only on an ordinal scale. AHS provides limited data on assets—ordinal indicators of “whether savings are present,” investments in a farm or business, the “dollar amount of savings or investments over \$20,000,” and the value of owner-occupied homes. Although some analysts have devised ways of estimating other assets from AHS (see Quercia et al. 1998), the requisite assumptions render such figures incomparable with variables available in other surveys. SIPP reports interval-scale data for both people and households by several asset categories (e.g., savings bonds, savings and checking accounts, and the net value of stocks). CES also offers interval measures for a detailed set of assets, as does PSID (see table A.2 for details). Therefore, housing researchers interested in asset information might consider SIPP, CES, and PSID. When selecting from among these three, researchers should evaluate the relative importance of specific asset information in relation to sample size, geographic scale of reporting, and other factors.

Finally, information on house value varies widely among these data sets. CPS provides no information, whereas SFs give only aggregate figures. PUMS, by contrast, gives a fairly detailed breakdown of ordinal categories, whereas a more coarse breakdown is found in AHS and NSFH. CES and PSID offer interval, self-reported estimates of home values, and SIPP gives an estimate inferred from responses to separate questions on equity and mortgage balance.

Empirical Analysis of the Data Sets

What do the data sets reveal about the socioeconomic characteristics of renters and owners? Do these characteristics vary systematically across different data sets? We consider these issues for a subset of the data sources: AHS, CES, NSFH, and SIPP. The data accessed from these four sources were collected at a comparable period of time, from the early to mid-1990s.

⁷ The periodic debt payment is considered in the allowable “back-end ratio” guideline (total monthly debt obligations as a share of gross monthly income) of different mortgage instruments.

1. AHS: 1995 AHS
2. CES: Consumer unit (CU) characteristics and income file for the fourth quarter of 1993
3. NSFH: Wave 2 respondent file, conducted between 1992 and 1994
4. SIPP: 1993 panel wave 7, core and topical module microdata files (wave 7 consists of four rotation groups, each interviewed in different months from October 1994 to January 1995)

Because we want to compare information across the data sets, we focus on variables that are defined consistently in the various information resources. That variable comparability is straightforward for characteristics such as household size and income, but there is considerable variation in the reporting of assets. For example, not all of the data sets report business and house equity. To be consistent, we focus on the asset value that is *net* of business and house equity.⁸

To determine if differences in mean values across data sets are statistically significant, we used the Wald test, which is described in greater detail in appendix B.⁹ This approach tests the significance of the difference between two means based on independent samples. The Wald test simply allows one to assess the possibility that sampling variability might explain any observed differences across the data sets.¹⁰

In addition to sampling errors, observed differences in estimates across different surveys are due to a multitude of other factors, including nonsampling errors and differences in survey timing, definitions of variables and units of analysis, and question wording. The nonsampling error results from undercoverage of the population, nonresponse on the part of the sample members, and errors made during data collection and processing (e.g., differences in the interpretation of questions, inability or unwillingness of the respondents to provide correct information, and errors in estimating values for missing data). It is worth noting that nonsampling errors can be particularly problematic for topics that people consider sensitive or that are difficult to recall with precision (e.g., annual income, value of assets, and property taxes). Quality control and edit procedures are used to reduce errors made by respondents, coders, and interviewers, but a significant part of the nonsampling error usually remains unaccounted for.

⁸ To simplify the presentation and discussion, this article includes only selected values to highlight relevant points. More detailed information is available from the authors on request, including information (mean, median, standard error of the mean, and minimum-maximum values) derived from AHS, CES, NSFH, and SIPP on the socioeconomic characteristics of renters and owners by race and Hispanic origin. The specific procedures used to select observations, develop comparable estimates of household assets, and calculate standard errors also are available on request.

⁹ We could not apply the Wald test to values derived from NSFH because we could not calculate NSFH's standard errors. NSFH documentation includes neither pseudohalfsample and pseudostratum codes (such as those included in SIPP and CES data sets) used in pseudo replication techniques, nor formulas and parameters (such as those provided in SIPP and AHS documentation) used in generalized standard error calculations. Our study of the analysis of variation across the data sets therefore is limited to AHS, CES, and SIPP.

¹⁰ In interpreting the Wald test results, one also should be aware that estimates of sampling errors themselves are subject to sampling error and, often, to biases as well (see, for example, U.S. Bureau of the Census [1996] for possible sources of bias in sampling error estimates for SIPP).

Any differences in estimates across surveys that are large enough to be beyond the likely bounds of sampling variability reflect the net effect of the other factors mentioned above. It is beyond the scope of this article—and difficult for practical purposes—to examine the separate contributions of these other factors to the differences in estimates across surveys and, relatedly, to provide guidance as to which data sources provide the “correct” estimates for variables that vary across surveys.

A Snapshot of Renters and Owners

Table 1 presents median values for selected socioeconomic and housing characteristics for owners and renters. A simple comparison of the values indicates that households look quite different when viewed through the lens of different surveys. Although household size and the age of survey respondent show little variation across the data sets, the financial profile of both renters and owners varies considerably. In several cases the disparities are striking. The median renter surveyed in CES has no assets, but NSFH and SIPP suggest median values of \$750 and \$400, respectively. For owners, CES implies no household assets (excluding, as noted, business and house equity), whereas progressively higher estimates of the median value of assets are obtained from SIPP (\$4,654) and NSFH (\$15,000). Household incomes also vary substantially across the data sources. For renters, incomes range from less than \$14,000 in CES to nearly \$24,000 in NSFH. For owners, median income is about \$27,000 in CES, but it is \$43,000 in NSFH.

Our comparison also reveals differences in some housing characteristics across sources. Monthly rent is about 25 percent higher in AHS (\$472) than in NSFH (\$375) or SIPP (\$379).¹¹ Estimates of housing expenditures as a proportion of income also fluctuate substantially, whether measured at the aggregate or household level. The median renter in AHS spends 23.1 percent of gross income on rent, compared with only 17.5 percent for the NSFH sample and 19.4 percent for the SIPP sample.

These results raise more questions than they answer. They confirm the numerous empirical studies that show differences in the socioeconomic characteristics of owners and renters. However, these results also suggest that studies incorporating both owners and renters are likely to be quite sensitive to the distinctive characteristics of particular surveys chosen by the researchers.

Results from Wald Tests

Are the differences in household and housing characteristics across data sets statistically significant? Consider the most basic financial indicator—household income. Results from the application of the Wald test suggest that differences in mean household income estimates from AHS, CES, and SIPP are statistically significant (table 2). Regardless of tenure, CES

¹¹ Rent is reported by the respondent and is typically the rent paid by the respondent to the landlord. Questions related to rent vary somewhat across the surveys, potentially creating inconsistencies in the inclusion of items such as government subsidies in the reported amount.

Table 1. Socioeconomic Profile of Renters and Owners, Median Values

Characteristic	AHS	CES	NSFH	SIPP
Renters				
Household assets (\$)	NA	0	750	400
Household income (\$)	23,200	13,900	23,800	20,997
Household size	2.0	2.0	2.0	2.0
Age of reference person (years)	37.0	34.0	38.0	37.0
Monthly rent (\$) ^a	472	NA	375	379
Rent-to-income ratio (%)				
Aggregated ^b	24.4	NA	18.9	21.7
Household level ^c	23.1	NA	17.5	19.4
Sample size	14,731	1,999	2,719	5,936
White-black ratios				
Income	1.60	1.55	1.62	1.51
Assets	NA	NA	NA	116.67
Rent	1.19	NA	1.33	1.18
White-Hispanic ratios				
Income	1.25	1.12	1.30	1.28
Assets	NA	NA	1.00	14.00
Rent	0.95	NA	0.89	0.92
Owners				
Household assets (\$)	NA	0	15,000	4,654
Household income (\$)	42,500	27,030	43,000	38,004
Household size	2.0	2.0	3.0	2.0
Age of reference person (years)	49.0	50.0	48.0	50.0
House value (\$)	95,000	NA	89,000	85,000
Value-to-income ratio (%)				
Aggregated ^b	223.5	NA	207.0	223.7
Household level ^c	225.4	NA	200.0	221.7
Sample size	25,444	3,275	6,474	11,967
White-black ratios				
Income	1.47	1.57	1.45	1.34
Assets	NA	NA	25.33	9.67
House value	1.54	NA	1.73	1.55
White-Hispanic ratios				
Income	1.22	1.05	1.44	1.16
Assets	NA	NA	8.44	4.84
House value	1.05	NA	1.13	1.20

Notes: AHS = American Housing Survey, CES = Consumer Expenditure Survey, NA = Not available, NSFH = National Survey of Families and Households, SIPP = Survey of Income and Program Participation.

^a Rent is reported by the respondent and is typically the rent paid by the respondent to the landlord. Questions related to rent vary somewhat across the surveys, potentially creating inconsistencies in the inclusion of items such as government subsidies in the reported amount.

^b Calculated as the ratio of median annual rent (house value) to median income.

^c Calculated as the median value of household ratios between annual rent (house value) to household income.

income estimates are consistently lower than those in AHS. Mean income figures for both owners and renters in CES are about two-thirds of the values reported in AHS. CES income figures also are consistently lower than SIPP figures, although the gap is slightly smaller.

Table 2. Comparison of Socioeconomic Characteristics of Renters and Owners, Mean Values and Significance Tests

Characteristic	Mean Value			Wald Tests		
	AHS 1	CES 2	SIPP 3	Significant Differences across Data Sets at $p < 0.05$		
Renters						
Household assets (\$)	NA	792	8,164			2<3
Household income (\$)	27,834	18,745	26,915	1>2	1>3	2<3
Household size	2.4	2.2	2.4	1>2		2<3
Age of reference person (years)	41.0	39.1	42.2	1>2	1<3	2<3
Monthly rent (\$) ^a	495	NA	396		1>3	
Rent-to-income ratio (%) ^b	21.4	NA	17.7		1>3	
Sample size	14,731	1,999	5,936			
White-black ratios						
Income	1.50	1.29	1.39		1>3	
Assets	NA	3.17	7.59			2<3
Rent	1.21	NA	1.19			
White-Hispanic ratios						
Income	1.29	1.09	1.30			
Assets	NA	5.33	4.27			
Rent	0.98	NA	0.98			
Owners						
Household assets (\$)	NA	4,244	32,601			2<3
Household income (\$)	51,339	34,650	44,757	1>2	1>3	2<3
Household size	2.8	2.6	2.7	1>2	1>3	
Age of reference person (years)	51.8	52.6	52.0	1<2		2>3
House value (\$)	119,507	NA	104,712		1>3	
Value-to-income ratio (%) ^b	232.8	NA	234.0			
Sample size	25,444	3,275	11,967			
White-black ratios						
Income	1.37	1.35	1.32			
Assets	NA	2.04	6.96			2<3
House value	1.58	NA	1.58			
White-Hispanic ratios						
Income	1.18	1.05	1.18			
Assets	NA	4.62	3.10			
House value	1.10	NA	1.16			

Notes: AHS = American Housing Survey, CES = Consumer Expenditure Survey, NA = Not available, SIPP = Survey of Income and Program Participation.

^a Rent is reported by the respondent and is typically the rent paid by the respondent to the landlord. Questions related to rent vary somewhat across the surveys, potentially creating inconsistencies in the inclusion of items such as government subsidies in the reported amount.

^b Mean rent (value)-to-income ratio is computed as the ratio of mean annual rent (house value) and mean income.

The consistently lower income figures from CES might mostly be related to the definition of the unit of analysis in this survey—the CU. A CU is likely to be smaller than its counterpart (the household) in the other surveys. For example, in CES, three financially independent people sharing a house are considered three separate CUs (and their incomes are counted separately), but, according to the other surveys, they make up a single household (and their incomes are aggregated).

The situation is even more problematic for household assets. See Kennickell and Starr-McCluer (1997) for a discussion of the reliability of data regarding assets. Although the procedures used in an attempt to obtain comparable asset measures may be expected to introduce some nonsampling bias, the observed differences are far beyond any reasonable allowance for error. Asset measures exhibit astonishing levels of variation. The average asset value (excluding business and house equity) of renters is less than \$1,000 in CES, about one-tenth of the corresponding estimate from SIPP. The respective estimates for owners show a difference by a factor of eight. The Wald test indicates that the differences across data sets are significant for both renters and owners.

In contrast to the considerable variation in economic measures, we find greater correspondence in social and demographic characteristics such as age and household size. Overall, the Wald test results indicate fewer significant differences in these measures than for income and other economic variables. Still, AHS yields larger estimates of household size, and its estimates for the age of the reference person are older than CES estimates for renters but younger than CES estimates for owners (table 2). In general, alternative database selection is more likely to affect results if a researcher focuses on economic variables as opposed to basic demographic estimates.

One reason for the small differences in demographic profiles across various surveys might be the similarity of weighting schemes used for certain demographic variables. Each survey constructs sampling weights to ensure unbiased estimators for the characteristics being estimated in the full population. Weight computation is usually a multistage process that starts with computing a basic weight equal to the inverse of the probability of selection, followed by various postsampling adjustments to the basic weight. As a weight adjustment, some surveys align the sample distribution to the population distribution of sex, race/ethnicity, and age obtained from the decennial census or CPS (which itself is brought into agreement with the decennial census). As a result, it is expected that these demographic variables differ little across data sets with similar weighting schemes. The weighting schemes for AHS, NSFH, and SIPP all include adjustments based on sex, race/ethnicity, and age.¹² This may explain the relatively small differences in age statistics across surveys.

In light of the wide-ranging implications of entrenched racial and ethnic inequalities in housing markets, it is worth considering whether alternative data sets yield the same findings on intergroup differences in income and wealth. We find considerable sensitivity in this area (table 2). For example, measured by average household income, non-Hispanic whites consistently fare better than blacks and Hispanics, but the gap varies depending on the data set. For black and white renters, disparities in income are significantly greater in AHS than in SIPP, whereas no significant difference in disparities is found between AHS and CES. Among black and white homeowners, however, income disparities display no significant variation across the three data sets. The ratio of white to Hispanic income is not significantly different across the three data sets, regardless of tenure.

¹² Based on the available survey documentation, it is unclear whether CES includes a similar adjustment.

Racial-ethnic differences in average household assets are much more pronounced than income differences (table 2). However, the magnitude of the gap varies between CES and SIPP, depending on the groups being compared. For blacks and whites, disparities are greater when measured in SIPP (7.6 to 1 for renters, 7.0 to 1 for owners). However, for whites and Hispanics, the disparities are not significantly different between CES (5.3 to 1 for renters, 4.6 to 1 for owners) and SIPP (4.3 to 1 for renters, 3.1 to 1 for owners).

Conclusion

In recent years, analysts have turned to an ever-widening range of data sources to measure the dynamics of America's housing. Yet surprisingly little systematic comparison of household and housing characteristics across alternative data sets has been made. Our limited analysis of some of the more popular sources indicates that demographic characteristics generally are comparable across data sets, but numerous, statistically significant differences exist in estimates for the most basic household financial measures. These variations represent the net effect of innumerable factors, including differences in respondent selection bias, panel and cross-sectional design considerations, specific interview protocols, question wording, definitions of household, and sample weighting procedures.

There is no "correct" database for housing research, and in most cases it even is difficult to identify a preferred source without considering the precise analytical requirement of a specific empirical question. Surveys are imperfect abstractions of the social world, and professionals involved in survey design know only too well that context is paramount. How someone estimates his or her household income, for example, depends in part on the scope, detail, and length of all the other questions asked in the interview. Indeed, this most basic measure of household financial circumstances varies considerably across different surveys. Alternative sources yield dramatically varied estimates of what are—in all mainstream theories—the same things.¹³ These different estimates find their way into many different types of policy benchmarks used by federal, state, and local agencies—and have effects (however small) on the housing market realities that future social surveys will seek to measure. Ultimately, it is impossible to predict all the consequences of measurement differences across social surveys, but it is clear that analysts and policy makers should beware the illusion that data are known facts. Known facts are in short supply, but data sets offer a variety of starting points and reasonable assumptions.

¹³ It is crucial to note that our analysis has not even raised the possibility of deeper epistemological concerns regarding empirical measurement. Even if the technical considerations highlighted in our analysis could be controlled to eliminate cross-survey differences in, for example, house value, we still are left with a theory-laden observation. For instance, what kind of "known fact" would we have about a homeowner surveyed in the bubble economic climate of early 2000 who gives a wildly inflated estimate of what her home would sell for? Does the estimate acquire greater factual value if she is able to borrow against it in a home equity loan, or if she buys a new sport utility vehicle in accordance with the widely discussed wealth effect?

Appendix A
Database Profiles and Content

Table A.1. Core Housing Data Sources (Census, CPS, AHS)

Characteristic	Public Use		American Housing Survey (AHS)
	Summary Files (SFs) from the Decennial Census	Microdata Samples (PUMS) from the Decennial Census	
Profile information			
Produced by	U.S. Bureau of the Census	U.S. Bureau of the Census	U.S. Department of Housing and Urban Development and U.S. Bureau of the Census
General description/ purpose	Provides a census of the U.S. population and housing; data vary across SFs 1 and 2, which are based on the short forms, and SF 3 and 4, which are based on the long forms	Consists of microdata generated from the U.S. census; a 1 percent sample from the nation, and 5 percent samples for states and public use microdata areas (PUMAs)	Describes current housing conditions across the United States, including cost, quality, size, tenure, and financing
Most recent year available	2000 (SF 4 data available during 2003)	1990 data and earlier are available 2000 data available during 2003	1999
Periodicity	Every 10 years	Every 10 years	Every other year since 1981; metropolitan statistical area-specific survey every four years
Sample units	Housing units/households Individuals	Housing units/households Individuals	Housing units/households

Table A.1. Core Housing Data Sources (Census, CPS, AHS) (continued)

Characteristic	Public Use		American Housing Survey (AHS)
	Summary Files (SFs) from the Decennial Census	Microdata Samples (PUMS) from the Decennial Census	
Universe	All households and people in the United States	All people and households in the United States	All housing units, which includes apartments, single-family homes, mobile homes, and vacant housing units
Sample size	Variables in SFs 3 and 4 (derived from the long form) are based on a sample (generally 1 in 6 people or housing units) and are subject to sampling variability, especially at smaller scales	1 percent of the total U.S. population for the national characteristics file 5 percent of the total U.S. population for state files	About 50,000 households in 1999
Geographic extent	National	National	National
Geographic scale	National, regional, state, metropolitan statistical area, county, place, tract, block group, and blocks (for selected short-form tables)	National, state, super-PUMAs (1 percent file), and PUMAs (5 percent files)	National, regional, and metropolitan, including suburbs and central cities
Specific limitations			
Comments		A super-PUMA has a minimum population of 400,000; a PUMA has a minimum population of 100,000	Tends to underreport income and overreport poverty

Table A.1. Core Housing Data Sources (Census, CPS, AHS) (continued)

Characteristic	Public Use		American Housing Survey (AHS)
	Summary Files (SFs) from the Decennial Census	Microdata Samples (PUMS) from the Decennial Census	
Availability on World Wide Web	Extracts available at: http://www.census.gov	Microdata set available at: http://www.census.gov/des/11	Microdata set available at: http://www.census.gov/des/11
Content information—selected socioeconomic and housing data			
Social data			
Household size and composition	Household and family sizes and compositions Presence and age of children	Household and family sizes and compositions Age and relationship of members	Number of people in household Household type, chosen from 18 unique categories
Measurement scale for age of householder	Interval	Interval	Ordinal, with 8 categories
Race/ethnicity	Aggregate information In 1990, five major race categories and Hispanic/non-Hispanic ethnicity identification In 2000, six major race categories	Individual scale In 1990, five major race categories and Hispanic/non-Hispanic ethnicity identification In 2000, six major race categories	Five race categories and Hispanic/non-Hispanic ethnicity identification

Table A.1. Core Housing Data Sources (Census, CPS, AHS) (continued)

Characteristic	Public Use		American Housing Survey (AHS)
	Summary Files (SFs) from the Decennial Census	Microdata Samples (PUMS) from the Decennial Census	
Financial data			
Income	Ordinal scale Income from eight sources	Interval scale Income from eight sources, including wages and salary, farm and nonfarm self-employment, interest and dividends, Social Security, public assistance, retirement, and others	Household and family income reported in 15 categories Income from 12 sources reported as yes/no
	Assets	Value of specified owner-occupied homes but not home equity	Home equity (interval) Can calculate home equity
Liabilities	Mortgage status	Mortgage status and monthly mortgage payments	None
	Employment history	Previous year's labor participation Occupation/industry Self-employment status	Current occupational status and major industry Full-time/part-time status Self-employment status Employment stability
Credit history	None	None	None

Table A.1. Core Housing Data Sources (Census, CPS, AHS) (continued)

Characteristic	Summary Files (SFs) from the Decennial Census	Public Use Microdata Samples (PUMS) from the Decennial Census	Current Population Survey (CPS) Annual Demographic Survey	American Housing Survey (AHS)
Mortgage information	Household status as owner or renter Mortgage status Mortgage payments (not reported separately, but included in “selected monthly owner costs”)	Monthly mortgage payments, in interval scale up to \$1,999 Two variables report if payment includes taxes and insurance	None	Outstanding principal, in 16-level ordinal scale Year originated and terms Current interest rate Items included in mortgage payments Major source of down payment (from nine categories)
Property taxes	Yearly property taxes Ordinal scale with 15 categories	Yearly property taxes paid Ordinal scale with 60 categories	None	Average monthly real estate costs, in seven-level ordinal scale Annual taxes paid per \$1,000 in value
Housing data	Aggregate information by tract, among others Type of structure (categorical) Year structure built (ordinal scale) Unit characteristics such as number of rooms and bedrooms, plumbing facilities, and kitchen facilities	Detailed information on both owned and rented units, including number of rooms, number of bedrooms, and year built (in ordinal scale with eight categories) Type of structure (one-family attached, one-family detached, apartment, or three- to four-unit apartment, among others; 10 categories in total)	Type of living quarters (12 categories) Number of units in structure (five categories)	Information about the unit and the structure, including number of stories in structure, presence and condition of common stairways, external building conditions, foundations, water source, equipment available, air conditioning, heating, and plumbing

Table A.1. Core Housing Data Sources (Census, CPS, AHS) (continued)

Characteristic	Public Use		
	Summary Files (SFs) from the Decennial Census	Microdata Samples (PUMS) from the Decennial Census	Current Population Survey (CPS) Annual Demographic Survey
Neighborhood characteristics	The best source for neighborhood (i.e., census tract) descriptions: a wealth of socioeconomic variables, age and type of housing stock, and fine geographic scale Limited information available about physical characteristics of neighborhood housing (e.g., number of vacant units and plumbing facilities)	PUMAs are based on collections of census tracts; all census data can be aggregated to describe the characteristics of a PUMA The scale, 100,000 to 250,000 people, does not describe a neighborhood, but could be useful in larger cities	None
Tenure status	Owner or renter	Owner or renter	Owner or renter
House value	Aggregate information only Ordinal scale	Ordinal scale with 25 categories	None Ordinal scale with 15 categories
			Only source that asks about the "conditions surrounding your unit" and the conditions of other units on the block; unfortunately, the scale (suburb or city) is too coarse to describe a neighborhood
			American Housing Survey (AHS)

Table A.2. Specialized Surveys That Contain Housing Data (SIPP, CES, PSID)

Characteristic	Survey of Income and Program Participation (SIPP)	Consumer Expenditure Survey (CES)	Panel Study of Income Dynamics (PSID)
Profile information			
Produced by	U.S. Bureau of the Census	U.S. Department of Labor, Bureau of Labor Statistics	University of Michigan Survey Research Center
General description/ purpose	To collect current information on income, labor force participation, government program participation, and income distribution, and to estimate future costs of various federal, state, and local programs	To collect information on the consumption patterns of households, families, and individuals; follows a rotating panel of respondents who are interviewed once per quarter for five quarters and asked how they spent their money in the previous three months	To permit longitudinal analysis of family economic well-being and labor market participation, and to assess the intergenerational dimensions of work
Years available	1984 to 1993 and 1996	1979 to 1995 (older versions also are available)	Final release (Public Release II): 1968 to 1993 Early release (Public Release I): 1994 to 1997 for families 1994 to 1999 for individuals
Periodicity	Irregular	Quarterly from 1979 to present, and monthly from 1984 to present	Quarterly surveys over a 2.5-year period
Sample units	Individuals Households	A “consumer unit,” which approximates the household but also includes individuals in the household who are financially independent and make independent decisions	Individuals Families Households

Table A.2. Specialized Surveys That Contain Housing Data (SIPP, CES, PSID) (continued)

Characteristic	Survey of Income and Program Participation (SIPP)	Consumer Expenditure Survey (CES)	Panel Study of Income Dynamics (PSID)
Universe	Noninstitutional civilian population	Total noninstitutionalized population (urban and rural) of the United States	Nationally representative of U.S. individuals (men, women, and children) and family units
Sample size	For 1984 to 1993, approximately 15,000 to 20,000 households per year; for 1996, approximately 36,700 households	Approximately 10,000 interviews per quarter covering 5,000 consumer units	Gradually grown from 4,800 families in 1968 to more than 7,000 in the 2001 wave of data collection
Geographic extent	National	National	National
Geographic scale	States available, some metropolitan statistical area codes Inside and outside metropolitan statistical area and central city	Data available for the four census regions	National
Specific limitations	No surveys started for 1994 or 1995	Documentation suggests that data for income, assets, and liabilities may not be as accurate as the expenditure numbers	Early release data have not been cleaned; only experienced users should attempt to work with it
Comments	Consists of a core data set collected once every four months, supplemented by specialized "topical modules" that vary from wave to wave	Consists of a quarterly interview designed to collect information on large, infrequent purchases and a weekly diary designed to capture small purchases	Provides some support for work with final release data but not for early release data
Availability on World Wide Web	Microdata available at: http://www.census.gov/des/11	Microdata available at: http://www.census.gov/des/11	All years available at: http://www.isr.umich.edu/src/psid/

Table A.2. Specialized Surveys That Contain Housing Data (SIPP, CES, PSID) (continued)

Characteristic	Survey of Income and Program Participation (SIPP)	Consumer Expenditure Survey (CES)	Panel Study of Income Dynamics (PSID)
Content information—selected socioeconomic and housing data			
Social data			
Household size and composition	Family composition Age and presence of children Possible to construct subfamilies	Family composition Age and presence of children Possible to construct subfamilies	Family composition Age and presence of children Possible to construct subfamilies
Measurement scale for age of householder	Interval	Interval	Interval
Race/ethnicity	Five race categories, Hispanic or non-Hispanic identification	Race (black or white), Hispanic or non-Hispanic identification	Five race categories, Hispanic or non-Hispanic identification
Financial data			
Income	Interval scale Income from 38 sources Can calculate individual, household, and family income For annual estimates, can tally monthly totals for the three four-month waves	Interval scale Income from nine sources Income by household member	Interval estimates for the following: gross earnings; bonuses, tips, and overtime; income from boarders; interest, dividends, and trust funds; Aid to Families with Dependent Children, supplemental security income, Social Security, and other welfare; pension benefits; unemployment compensation; child support and alimony; and income from other household members

Table A.2. Specialized Surveys That Contain Housing Data (SIPP, CES, PSID) (continued)

Characteristic	Survey of Income and Program Participation (SIPP)	Consumer Expenditure Survey (CES)	Panel Study of Income Dynamics (PSID)
Assets	<p>Topical Module for waves 4 and 7 is "Assets and Liabilities"</p> <p>Interval scale per person for current savings in deposit accounts, savings bonds, insurance, and retirement accounts</p> <p>Equity in home, business, and motor vehicles</p>	<p>Interval scale</p> <p>Amount the consumer unit has in savings and checking accounts, estimated market value of stocks and bonds, dollar value of owned business, and insurance policies</p>	<p>For 1994, includes a supplemental set of questions on assets and wealth that provides interval estimates on a range of financial assets</p>
Liabilities	<p>Interval scale</p> <p>Amounts owed for store bills and credit cards, bank and credit loans, and "other debts," including medical bills</p> <p>Home mortgage (see below)</p>	<p>Interval scale</p> <p>Amounts owed to banks for finance charges, late payments, and interest charges; to insurance companies and credit card companies; and for medical bills</p> <p>"Total owed this month" and "total owed this month last year"</p> <p>Home mortgage (see below)</p>	<p>1994 wealth supplement includes some questions on outstanding debts such as home mortgage, auto loans, credit card debt, and other liabilities</p>
Employment history	<p>Average hours worked each week for the previous 30 months</p> <p>Occupation and industry</p> <p>Self-employment status</p>	<p>Employment history for the previous 12 months</p> <p>Full-time/part-time status</p> <p>Spouse's employment for the previous 12 months</p>	<p>Previous 12 months of employment history</p> <p>Average hours worked</p> <p>Self-employment status</p>
Credit history	None	None	Includes questions on "fiscal stress" and money management

Table A.2. Specialized Surveys That Contain Housing Data (SIPP, CES, PSID) (continued)

Characteristic	Survey of Income and Program Participation (SIPP)	Consumer Expenditure Survey (CES)	Panel Study of Income Dynamics (PSID)
Mortgage information	Interval scale for how much is owed on the current house Year the mortgage began, number of years left on the mortgage, type (fixed or adjustable), and government insurance (Federal Housing Administration/Veterans Affairs)	Interval scale Price paid for house, size of down payment, and total closing costs Year acquired Current mortgage interest rate, length of mortgage, and monthly payments (principal and interest)	Number of years paid and years left to pay Type of mortgage Fixed or variable rate Government insured or not
Property taxes	None	Interval scale Federal, state, and local "personal property taxes"	None
Housing data			
Property characteristics	Type of structure (categorical)	Type of structure (single-family house, town house, high-rise, or multifamily house) Number of bedrooms Lot size (ordinal scale)	Type of structure (single-family, two-family house, or apartment) Number of rooms in unit
Neighborhood characteristics	None	None	None
Tenure status	Owner or renter	Owner or renter	Owner or renter
House value	Interval scale Estimated by combining equity in house and outstanding principal on mortgage	Interval scale Self-reported by respondent	Interval scale Self-reported by respondent

Table A.3. Specialized Surveys That Contain Housing Data (SCF, NSFH)

Characteristic	Survey of Consumer Finances (SCF)	National Survey of Families and Households (NSFH)
Profile information		
Produced by	Federal Reserve	University of Wisconsin
General description/ purpose	Designed to measure changes in income, net worth, assets, and liabilities for families in the United States, especially those in the top 10 percent of the income distribution	A two-wave survey designed to collect detailed information on families, family history, child-rearing, marital and parental arrangements, and children; includes information on housing
Years available	1983, 1986, 1989, 1992, 1995, 1998	1987 to 1988 (wave 1) 1992 to 1994 (wave 2)
Periodicity	Every three years	Not applicable
Sample units	"Families," which are defined similarly to households; for example, a one-person household is considered a family	Families Households Individuals (including parents and children)
Universe	National households with an oversampling of wealthy households	Nationally representative of the population
Sample size	1992: 3,906 respondents 1995: 4,299 respondents 1998: 4,309 respondents Includes wealthy households selected from high-earner tax returns from the Internal Revenue Service	Wave 1: 13,007 people, including a main cross-section of 9,637 people plus an oversampling of blacks, Puerto Ricans, Mexican Americans, single-parent families, families with stepchildren, cohabiting couples, and recently married people Wave 2: 10,008 of the people from wave 1
Geographic extent	National	National
Geographic scale	Nothing below the national scale	Nothing below the national scale
Specific limitations	Only 2,456 low- and moderate-income respondents in 1995	

Table A.3. Specialized Surveys That Contain Housing Data (SCF, NSFH) (continued)

Characteristic	Survey of Consumer Finances (SCF)	National Survey of Families and Households (NSFH)
Comments		
Availability on World Wide Web	Available at: http://www.federalreserve.gov/	Available at: http://www.ssc.wisc.edu/nsfh/home.htm
Content information—selected socioeconomic and housing data		
Social data		
Household size and composition	Household size Number of children	Family size Marital status Distinguishes between family and household
Measurement scale for age of householder	Interval, for head of the family	Interval age reported for each respondent
Race/ethnicity	White, black, Hispanic, and other for microdata set	Race and ethnicity reported, such as white, black, and Hispanic
Financial data		
Income	Income and “normal income” as continuous variables	Ordinal scale with 11 categories Income from seven sources, including wages and salaries, self-employment, Social Security, public assistance, child support or alimony, and interest or dividends Available for each individual

Table A.3. Specialized Surveys That Contain Housing Data (SCF, NSFH) (continued)

Characteristic	Survey of Consumer Finances (SCF)	National Survey of Families and Households (NSFH)
Assets	<p>Continuous variables</p> <p>Financial assets, including checking, savings, and money market deposit accounts; certificates of deposit; money market mutual funds; call accounts at brokerage houses; mutual funds; defined benefit pension accounts; bonds; and stocks</p> <p>Nonfinancial assets, including value of all vehicles, business vehicles, primary residence, other residential real estate, net business interests, and other miscellaneous nonfinancial assets</p>	<p>Ordinal scale with nine categories</p> <p>Savings and certificates of deposit, business and other investments</p>
Liabilities	<p>Mortgage on primary residence, debt on other residential property, credit card debt, installment debt, and "other" debt</p>	<p>Ordinal scale with seven categories</p> <p>Credit card and installment debts, educational loans, personal loans, motor vehicles, and other</p>
Employment history	<p>Current work status of householder</p> <p>Occupation (five categories)</p> <p>No employment history</p>	<p>Annual employment status from 1987 to 1992/1994 and if the same employer since 1987</p> <p>Occupational status</p>
Credit history	<p>Attitudes toward credit</p> <p>Whether respondent has been turned down for credit in the previous five years and why</p>	<p>Ordinal scale</p> <p>Amount in debt and if payments are overdue by more than two months</p>
Mortgage information	<p>Housing debt, including mortgages, home equity loans, and home equity lines of credit</p>	<p>Mortgage payments</p> <p>Ordinal scale with 11 categories</p>
Property taxes	<p>None</p>	<p>Property taxes paid in previous year</p> <p>Ordinal scale with eight categories</p>

Table A.3. Specialized Surveys That Contain Housing Data (SCF, NSFH) (continued)
 National Survey of Families and Households (NSFH)

Characteristic	Survey of Consumer Finances (SCF)	National Survey of Families and Households (NSFH)
Housing data		
Property characteristics	None	None
Neighborhood characteristics	None	None
Tenure status	Owner or renter	Owner or renter
House value	Median value from income categories	Self-reported Ordinal scale with nine categories

Appendix B

Wald Test Calculations

The Wald statistic tests for the significance of the differences between two means based on independent samples. Let \bar{X}_1 and \bar{X}_2 be the two weighted means of a given item derived from two different survey samples. Assuming that \bar{X}_1 and \bar{X}_2 normally are distributed with variances $V(\bar{X}_1)$ and $V(\bar{X}_2)$, and the underlying samples are independent, under the null hypothesis that the two estimates have the same expected value, the value $\bar{X}_1 - \bar{X}_2$ has a mean of zero and the variance $V(\bar{X}_1) + V(\bar{X}_2)$.

A test that the difference between the two means is zero can be based on the Wald statistic

$$W = (\bar{X}_1 - \bar{X}_2)^2 [V(\bar{X}_1) + V(\bar{X}_2)]^{-1},$$

which has a chi-squared distribution with one degree of freedom. If $W > 3.84$, the null hypothesis can be rejected at the 5 percent significance level. Table B.1 presents the detailed Wald results for selected comparisons of the samples across the data sets.

Table B.1. Univariate Wald Test for Mean Comparisons across Surveys

Group	Survey	Monthly Rent/ House Value		Household Assets		Household Income		Household Size		Age of Reference Person	
		AHS	CES	AHS	CES	AHS	CES	AHS	CES	AHS	CES
Renters	SIPP	414.71*	NA	NA	223.60*	5.66*	135.47*	0.39	17.12*	16.26*	31.12*
	AHS		NA		NA		181.22*		16.27*		13.55*
Owners	SIPP	188.38*	NA	NA	839.28*	218.40*	158.63*	20.68*	2.28	0.79	4.24*
	AHS		NA		NA		444.07*		16.06*		8.21*

Note: The values in the table represent the value of the Wald statistics. AHS = American Housing Survey, CES = Consumer Expenditure Survey, NA = Not available, SIPP = Survey of Income and Program Participation.

* $p \leq 0.05$.

References

- Boehm, Thomas P. 1995. A Comparison of the Determinants of Structural Quality between Manufactured Housing and Conventional Tenure Choices: Evidence from the American Housing Survey. *Journal of Housing Economics* 4(4):373–91.
- Boehm, Thomas P., and Alan M. Schlottmann. 2000. Housing and Wealth Accumulation: Intergenerational Impacts. Paper presented at Joint Center for Housing Studies of Harvard University Symposium on Low-Income Homeownership as an Asset-Building Strategy, November 14–15, Cambridge, MA.
- Calhoun, Charles A., and Marya T. Stark. 1997. Credit Quality and Housing Affordability of Renter Households. Paper presented at European Society for Population Economics Annual Meeting, June 12–15, Uppsala, Sweden, and at Southern Economic Association Annual Meeting, November 22–25, Washington, DC.
- Case, Bradford, Henry O. Pollakowski, and Susan M. Wachter. 1991. On Choosing among Housing Price Index Methodologies. *Journal of the American Real Estate and Urban Economics Association* 19:286–307.
- DiPasquale, Denise, and C. Tsurriel Somerville. 1995. Do House Price Indices Based on Transacting Units Represent the Entire Stock? Evidence from the American Housing Survey. *Journal of Housing Economics* 4(3):195–229.
- Duca, John V., and Stuart S. Rosenthal. 1994. Borrowing Constraints and Access to Owner-Occupied Housing. *Regional Science and Urban Economics* 24(3):301–22.
- Duda, Mark, and Eric S. Belsky. 2000. The Anatomy of the Low-Income Homeownership Boom in the 1990s. Paper presented at Joint Center for Housing Policy in the New Millennium Conference, October 2–3, Cambridge, MA.
- Early, Dirk W. 1998. The Role of Subsidized Housing in Reducing Homelessness: An Empirical Investigation Using Micro-Data. *Journal of Policy Analysis and Management* 17(4):687–96.
- Early, Dirk W., and Jon T. Phelps. 1999. Rent Regulations' Pricing Effect in the Uncontrolled Sector: An Empirical Investigation. *Journal of Housing Research* 10(2):267–85.
- Eller, T. J., and Wallace Fraser. 1995. *Asset Ownership of Households: 1993*. Washington, DC: U.S. Bureau of the Census.
- Galster, George, Laudan Aron, Peter Tatian, and Keith Wilson. 1996. *Estimating the Number, Characteristics, and Risk Profile of Potential Homeowners*. Washington, DC: Urban Institute.
- Green, Richard K. 2000. Home-Owning, Social Outcomes, Tenure Choice, and U.S. Housing Policy. Paper presented at Joint Center for Housing Studies of Harvard University Symposium on Low-Income Homeownership as an Asset-Building Strategy, November 14–15, Cambridge, MA.
- Green, Richard K., and Michelle J. White. 1997. Measuring the Benefits of Homeowning: Benefits to Children. *Journal of Urban Economics* 41:441–61.
- Haurin, Donald R., Toby L. Parcel, and R. Jean Haurin. 2002. Does Homeownership Affect Child Outcomes? *Real Estate Economics*, forthcoming.
- Juster, F. Thomas, and Kathleen A. Kuester. 1991. Differences in the Measurement of Wealth, Wealth Inequality, and Wealth Composition Obtained from Alternative U.S. Wealth Surveys. *Review of Income and Wealth* 37(1):33–62.

Kennickell, Arthur B., and Martha Starr-McCluer. 1997. Retrospective Reporting of Household Wealth: Evidence from the 1983–1989 Survey of Consumer Finances. *Journal of Business and Economic Statistics* 15:452–63.

Kiel, Katherine A., and Jeffrey E. Zabel. 1997. Evaluation of the Usefulness of the American Housing Survey for Creating House Price Indices. *Journal of Real Estate Finance and Economics* 14(January–March):189–202.

Lang, Robert E., James W. Hughes, and Karen A. Danielsen. 1997. Targeting the Suburban Urbanites: Marketing Central-City Housing. *Housing Policy Debate* 8(2):437–70.

Lexicon Publications. 1990. *The New Lexicon Webster's Encyclopedic Dictionary of the English Language*. New York.

Listokin, David, Elvin Wyly, Brian Schmitt, and Ioan Voicu. 2002. *The Potential and Limitation of Mortgage Innovation in Fostering Homeownership in the United States*. Washington, DC: Fannie Mae Foundation.

Moore, Dora J. 1991. Forecasting the Probability of Homeownership: A Cross-Sectional Regression Analysis. *Journal of Housing Research* 2(2):125–43.

Nelson, Walt A. 1994. Demonstrating the Use of American Housing Survey Data in Cross-Sectional Studies. *Journal of Real Estate Literature* 2(1):69–77.

Pitkin, John R. 1998. Changes in Homeownership and Households, 1993 to 1995: An Evaluation of Estimates from the Current Population Survey. *Journal of Housing Research* 9(2):317–26.

Pollakowski, Henry O. 1995. Data Sources for Measuring House Price Changes. *Journal of Housing Research* 6(3):377–87.

Quercia, Robert, George McCarthy, Jerome Levy, and Susan Wachter. 1998. The Impacts of Affordable Lending Efforts on Homeownership Rates. Paper presented at annual midyear meeting of the American Real Estate and Urban Economics Association, May 26, Washington, DC.

Rasmussen, David W., Isaac F. Megbolugbe, and Barbara A. Morgan. 1995. Using the 1990 Public Use Microdata Sample To Estimate Potential Demand for Reverse Mortgage Products. *Journal of Housing Research* 6(1):1–23.

Rosenthal, Stuart S. 2000. Eliminating Credit Barriers To Increase Homeownership: How Far Can We Go? Paper presented at Joint Center for Housing Studies of Harvard University Symposium on Low-Income Homeownership as an Asset-Building Strategy, November 14–15, Cambridge, MA.

Savage, Howard A. 1997. *Who Can Afford To Buy a House in 1993?* Current Housing Reports H121/97–1. Washington, DC: U.S. Bureau of the Census.

Savage, Howard A. 1999. *Who Can Afford To Buy a House in 1995?* Current Housing Reports H121/99–1. Washington, DC: U.S. Bureau of the Census.

Savage, Howard A., and Peter J. Fronczek. 1993. *Who Can Afford To Buy a House in 1991?* Current Housing Reports H121/93–3. Washington, DC: U.S. Bureau of the Census.

Stegman, Michael A., Roberto Quercia, George W. McCarthy, Michael Foster, and William Rohe. 1991. Designing Better Homeownership Assistance Programs Using the Panel Study of Income Dynamics (PSID): An Exploratory Analysis. *Journal of Housing Research* 2(1):39–85.

U.S. Bureau of the Census. 1996. 1993 Survey of Income and Program Participation. Machine-readable data file and technical documentation.

Wieand, Kenneth F. 1983. The Performance of Annual Housing Survey Quality Measures in Explaining Dwelling Rentals in 20 Metropolitan Areas. *Journal of the American Real Estate and Urban Economics Association* 11(1):45-68.

