

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, the Office of Policy Development and Research 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.

Data Sources for U.S. Housing Research, Part 2: Private Sources, Administrative Records, and Future Directions

Daniel H. Weinberg
Virginia Tech

This article is the second of a two-part article about data sources for U.S. housing research. The first part, which appeared in the previous issue of Cityscape (Volume 16, Number 3), addressed public sources.

Abstract

For practitioners and policymakers to make a serious attempt to affect housing policy, they must cite evidence-based research. Part 2 of this article summarizes many of the private sources of housing data for researchers that can provide such evidence. It then summarizes the challenges of using administrative records (AR) and proposes to construct new data sources by marrying survey data with AR and by constructing synthetic databases. The article concludes with a brief discussion of some data issues.

Introduction

The basis for good housing policy is evidence-based research, and the only way to do good research on housing is to base that research on appropriate data. Whereas part 1 of this article focused on government data sources for U.S. housing statistics, part 2 describes private data sources and administrative records (AR). It concludes with suggestions for future data production activities and mentions two unresolved data issues.

Private-Sector Data Sources

The National Association of Realtors® (NAR) and the National Association of Home Builders (NAHB) both issue housing affordability indexes (the latter is known as the NAHB/Wells Fargo Housing Opportunity Index). Such an index typically indicates whether a family with median income can afford the median-priced existing single-family home at prevailing mortgage rates (NAR uses the national median income and NAHB uses the U.S. Department of Housing and Urban Development's [HUD's] Area Median Incomes).

National Association of Realtors

NAR also provides monthly data series that track housing market sales: monthly sales volumes for existing homes by region, monthly sales volumes for single-family and cooperative apartments, monthly sales inventories of existing single-family and condominium homes, and the monthly pending home sales index (a forecast of existing home sales in the subsequent 1 to 2 months).

National Association of Home Builders

On a subscription basis, NAHB also offers 43 sets comprising various data series of interest to its constituency. These sets include data on building material prices (for example, framing lumber), employment, and permits. NAHB surveys multifamily developers and property managers to produce a Multifamily Production Index and a Multifamily Vacancy Index.

Mortgage Bankers Association

For subscribers, the Mortgage Bankers Association Weekly Applications Survey offers a comprehensive analysis of mortgage application activity. Historical index data are available back to the original start date of the survey in 1990. The survey's 15 indexes cover fixed-rate, adjustable-rate, conventional, and government loans for purchases and refinances.

RealtyTrac

RealtyTrac® Inc. has a website with foreclosure listings covering more than 2 million default, auction, short-sale, and bank-owned homes. Access is available for subscribers, and bulk downloads can be licensed. RealtyTrac describes its data as covering more than 100 million homes in 2,200 counties, accounting for 85 percent of all properties in the largest 200 metropolitan areas in the United States. For each property, RealtyTrac provides detailed housing characteristics (equity,

foreclosure details, comparable sales and listings, trends, lot size, square footage, price, and year built) and sales history (historical loan positioning, loan-to-value ratio, loan amount, estimated market value, property information, default amount, owner name, trustee, and lender name). HUD and the U.S. Census Bureau began joint research using these records matched to the American Housing Survey (AHS). RealtyTrac is distinct from other companies providing online foreclosure information in that it has historical data back to 2005 and provides analytic reports.

Other Sources of Information

Companies such as CoreLogic, Inc., Black Knight Financial Services, Inc., and Zillow can provide basic property tax data, including parcel boundaries, for a large subset of the United States. As CoreLogic notes—¹

CoreLogic is the nation's largest provider of advanced property and ownership information, analytics and solutions. Our databases cover more than 99 percent of the U.S. properties. CoreLogic obtains property records, tax assessments, property characteristics, and parcel maps from tax assessors and county recorders offices across the nation. This information is combined with flood, demographics, crime, site inspection neighborhood, document images and other information from proprietary sources to further enrich our databases.

Zillow provides property-level data, including historical sales price and year, taxes, and number of bedrooms and bathrooms; demographic data at the city and neighborhood level; and neighborhood information, including the Zillow Home Value Index, median single-family home and condominium values, and average tax rates. The Census Bureau is looking into whether such commercial data sources can reduce the cost of updating the Master Address File (MAF).

MPF Research analyzes the rental housing market for clients. The company advertises that, “With exclusive access to a completely unique data source and a solid foundation of sound statistical methodologies, MPF Research publishes 72 individual apartment market reports covering the top 100 markets nationally.”² MPF Research presents little about its methodology on its website.³

The National Council of Real Estate Investment Fiduciaries (NCREIF) is a cooperative organization that publishes information provided by its members. Its website indicates that, “NCREIF was established to serve the institutional real estate investment community as a non-partisan collector, processor, validator and disseminator of real estate performance information.”⁴ NCREIF bases its reports on its database of all-equity properties begun in 1977. In 2013, NCREIF has information on “approximately 30,000 properties historically, and approximately 10,000 current properties. NCREIF collects 67 data fields each quarter that consist of financial information such as Market

¹ Quoted from <http://www.corelogic.com/solutions/property-information-analytic-solutions.aspx>.

² Quoted from <https://www.realtor.com/mpf-research/?src=AdWords&medium=PPC&campaign=AdGroupName&Network=Search&kw=mpf&gclid=CNfo5ayInsMCFdgKgQodtj0AmQ>.

³ They note at <https://www.realtor.com/mpf-research/methodology/> that “Data collected in the MPF Research quarterly survey is collected through various sources. Where available, MPF Research can incorporate data from RealPage software products. MPF Research also collects data through direct relationships with management companies, through telephone surveys, and through e-mail surveys that are completed by apartment community owners or managers.”

⁴ Quoted from <https://www.ncreif.org/about.aspx>.

Value, [Net Operating Income], Debt, and [Capital Expenditures], as well as descriptor data such as Property Type and Subtype, Number of Floors, Square Footage, Number of Units, and Location” (NCREIF, 2013–2014: 2). It also publishes the NCREIF Property Index (NPI), “which is a quarterly index tracking the performance of core institutional property markets in the U.S.” (NCREIF, 2013–2014: 2), and the NCREIF Transaction-Based Index (NTBI)—“The NTBI is an equal-weighted transaction and appraisal index while the NPI is a value weighted index calculated using appraised values” (NCREIF, 2013–2014: 7). Using a subset of the included properties, NCREIF also publishes the NCREIF Timberland and Farmland Property Indices and provides other products to its clients such as a quarterly property index trends analysis report and operations data categorized by “four subcategories within income and eight categories within expense, as well as four capital expenditures subgroups” (NCREIF, 2013–2014: 3).

The Promise and Challenge of Administrative Records

Researchers have probably reached the limits of what government survey data collection can accomplish. As federal budgets get tighter, fewer surveys (and reduced sample sizes) are likelier than increased coverage of topics or additional samples to provide separate information for more metropolitan areas. That likelihood suggests that a fruitful area for federal statistical agencies to create value added is to take advantage of existing AR data sources; that is, data collected primarily or exclusively for administrative purposes rather than for research, also known as *third party data*. As the HUD *Research Roadmap* points out—

Research has shifted toward administrative data for three reasons:

1. Administrative records offer much larger sample sizes for full populations, which support more compelling research designs and research into important but relatively rare events.
2. Administrative files often have an inherent longitudinal structure that enables researchers to follow individuals over time and address policy questions.
3. Administrative data are less likely than survey data to suffer from high and rising rates of nonresponse, attrition, and underreporting. (HUD, 2013: 3)

Harnessing the power of these data through web-based information systems and geospatial analysis and matching these data with survey and administrative data from other agencies will provide the foundation for the next generation of evidence-based policymaking.

One particularly important area for investigation is the use of AR for improving the Census Bureau MAF. Improving the MAF—the basis for all Census Bureau household survey samples—will yield benefits to all such surveys and to the next decennial census. Under an agreement with the Census Bureau, the U.S. Postal Service (USPS) already provides a copy of its Delivery Sequence File (DSF) twice a year, and each DSF is used to update the MAF. The Census Bureau is investigating the use of National Change of Address files for improving the MAF.

Two other key components of MAF updates associated with the decennial census—address canvassing to determine ground truth and local updates—can be brought further into the digital

age. Efforts under way through the Census Bureau Geographic Support Systems Initiative will establish links to counties and large cities that can provide periodic electronic updates to their address files. The Census Bureau currently has no plans to run the Local Update of Census Addresses program as an ongoing program rather than a once-a-decade program. True partnership between the Census Bureau and state and local governments to improve the address list should be a two-way street.

Through an interagency agreement with USPS, HUD receives counts of total and vacant business and residential addresses in the United States at the ZIP+4 geographic level. HUD uses these data for a variety of purposes, including researching neighborhood change, tracking disaster recovery, gauging the foreclosure crisis, analyzing housing markets, and measuring the effect of HUD funding on communities. HUD also makes the vacancy data available at the census tract level to government and nonprofit organizations through a permitted-user sublicense agreement.

HUD collects information on the tenants in HUD-subsidized housing in its Public and Indian Housing Information Center (PIC) system and its Tenant Rental Assistance Certification System (TRACS). Local program administrators use form HUD-50058 to submit data to the PIC system and form HUD-50059 to provide HUD with tenant data for TRACS. PIC data contain longitudinal information on families living in public housing or receiving tenant-based housing vouchers, whereas TRACS data contain longitudinal information on families living in project-based Section 8 housing. HUD uses these data in several ways and provides them for research purposes to other government agencies that promise confidentiality protection.⁵ Mast provides the following information about PIC.

The PIC system has quarterly entries for each family receiving HUD rental assistance starting in 1995. Data are available on income, rent, and a large number of other household and PHA [public housing agency] characteristics. ...

The PIC data system is transaction based. The most common transactions are (1) admissions, (2) annual [reexaminations], (3) interim [reexaminations] due to changes in eligibility factors such as income or family size, (4) moves, and (5) exits from the program. The system captures the most recent transaction at the end of each quarter. If multiple transactions for a household occur during a quarter, only the most recent is available. If no transaction occurs during a quarter, the family's entry is a duplicate of the entry for the previous quarter.

Rent contracts are effective for 1 year and most households have only one transaction per year. Therefore, most changes are made annually, not quarterly. (Mast, 2012: 60)

The HUD Office of Policy Development and Research produces annual tabulations from the PIC/TRACS data called Picture of Subsidized Households (the most recent is for 2009). As the website notes, "Picture of Subsidized Households describes the nearly 5 million households living in HUD-subsidized housing in the United States [providing] characteristics of assisted housing units and residents, summarized at the national, state, public housing agency (PHA), project, census tract, county, Core-Based Statistical Area and city levels as downloadable files."⁶ A 5-percent sample of the

⁵ For examples of research using the HUD-PIC extract file, see Lubell, Shroder, and Steffen (2003); Mills et al. (2006); Olsen et al. (2005); Shroder (2002); and Tatian and Snow (2005).

⁶ Quoted from <http://www.huduser.org/portal/datasets/picture/yearlydata.html#download-tab>.

microdata is available to qualified researchers. In addition, as mentioned previously, the 2011 AHS collected data from a supplementary sample of HUD-subsidized units selected from PIC/TRACS.

The Federal Financial Institutions Examination Council (FFIEC) collects data from lending institutions related to the enforcement of mortgage regulations. The Home Mortgage Disclosure Act (HMDA) was enacted by Congress in 1975 and was implemented by Federal Reserve Board Regulation C. On July 21, 2011, the rule-writing authority of Regulation C was transferred to the Consumer Financial Protection Bureau. Regulation C requires lending institutions to report public loan data to assist in—

- Determining whether financial institutions are serving the housing needs of their communities.
- Siting local public-sector investments so as to attract private investment to areas where it is needed.
- Identifying possible discriminatory lending patterns.

HMDA initially required reporting of the geographic location of originated and purchased home loans. In 1989, Congress expanded HMDA data to include information about denied home loan applications and the race, sex, and income of applicants and borrowers. In 2002, the Federal Reserve Board amended the HMDA regulations to require lenders to report price data for certain higher priced home mortgage loans and other new data. For each transaction, with some exceptions, the lender reports data about—

- The loan (or application), such as the type and amount of the loan made (or applied for) and, in limited circumstances, its price.
- The disposition of the application, such as whether it was denied or resulted in a loan origination.
- The property to which the loan relates, such as its type (single-family or multifamily) and location (including the census tract).
- The applicant's ethnicity, race, gender, and income.
- The sale of the loan (if applicable).

This regulation applies to certain financial institutions, including banks, savings associations, credit unions, and other mortgage-lending institutions. FFIEC also collects similar data from private mortgage insurance companies on a voluntary basis and is responsible for administering the regulations to implement the *Community Reinvestment Act of 1977*,⁷ “intended to encourage depository institutions to help meet the credit needs of the communities in which they operate.”⁸

Several countries maintain housing registers—a list of all housing units and their characteristics—that can form the basis for housing analysis. For example, Denmark established its first housing register in the 1880s for the city of Copenhagen. As Christensen noted—

The [Danish] Building and Housing Register (BBR) was established in 1977. Since 1981, BBR has been updated annually by the municipalities. Before 1981, data on housing

⁷ Public Law 95–128.

⁸ Quoted from <http://www.ffiec.gov/cra/default.htm>.

conditions were collected as part of nationwide census of all households in Denmark that took place every fifth year. The first nationwide census including housing information took place in 1955. BBR consists of national data concerning building and housing. The purpose of the register is to describe the total housing stock and individuals' housing conditions and is used for administrative purposes. ...

There are good opportunities to carry out research on Danish housing conditions. The key data in BBR are of high quality and go back in time so longitudinal analyses can be executed. Furthermore, BBR can be matched with other registers so it is possible to make detailed analyses of tenant composition over time. In particular, analyses that compare individuals over time living in different segments of the housing market, e.g. ownership, social housing sector, and private sector, provide unique knowledge of individuals' living and housing conditions. (Christensen, 2011: 106, 108)

No U.S. housing register exists, however. The closest approximation is MAF, which is confidential under federal law.⁹ Under Title 13, however, MAF can be accessed for research that also benefits the Census Bureau (through its network of Research Data Centers). The MAF contains little information other than the address and associated census geography, but it can be linked to many Census Bureau household surveys.

The public property records in the United States that are the basis for property taxes are also potential data sources. Because these records are assembled at the municipal level, however, they are of varying quality, such as might result from delays in reassessment. Companies such as Zillow aggregate these records to offer services to the public for specific addresses. Researchers may be able to access these records for their own research.

Promising Techniques for Creating Additional Data Sources

While the data in AR datasets are interesting and useful, their value can be enhanced for research purposes.

Linking

One method that can enhance the value of existing data is to link datasets together. In this section, I describe a recent effort (Andersson et al., 2013, in which I participated) that linked together decennial census data, unemployment insurance AR on earnings, and HUD administrative data on subsidies to create a new database for housing research.

Andersson et al. (2013) and ongoing research focus on a difficult research issue—analyzing how children's housing affects their earnings in early adulthood. Andersson et al. developed a frame of households and children from the internal version of the 2000 decennial census. The short form provided a set of demographic variables that can be used to control for observable characteristics of parents and children. It also provided the residential location of households in 2000, which Andersson et al. linked to neighborhood characteristic variables (aggregates of the long-form data

⁹ *Code of Federal Regulations*, Title 13.

to the block group and census tract levels). Next, they used person identifiers developed at the Census Bureau to link the parents and children to HUD-PIC, the administrative data file of housing assistance recipients described previously. The HUD-PIC file covered 1997 through 2005; it was used to identify each year a parent or child was in subsidized housing and whether they were in public housing or received a housing voucher enabling them to live in private-sector housing.

Finally, Andersson et al. used the unique person identifiers to link the children in the sample to earnings records for 2008 through 2010 (and parents to their income for the entire period). The Census Bureau Longitudinal Employer-Household Dynamics (LEHD) dataset provides earnings records for more than 130 million workers each quarter from the mid-2000s onward.¹⁰ Those records provided a measure of labor market outcomes for 1.8 million children who were ages 13 to 18 in 2000 in low-income families—a sample size sufficient to present results disaggregated by race and Hispanic origin, gender, and housing subsidy program, while controlling for neighborhood conditions such as poverty level. When the initial analysis is complete, analysis with the file can be expanded to other topics, such as residential mobility and intergenerational earnings mobility.

Synthetic Data

One key problem with using the American *Housing Survey* for housing analysis is the relatively small sample sizes in any one location (metropolitan area), though the sample sizes appear adequate for national analysis. One key problem with using the American *Community Survey* (ACS) for housing analysis is the relatively few questions asked about housing and neighborhood physical, social, and economic characteristics. Is there any way to combine the strengths of the two surveys to enhance the data available for housing analysis?

Recent work by Reiter and others suggests it is possible to create a (partially) synthetic dataset that combines AHS and ACS using exact matches and modeling.¹¹ Synthetic datasets are created based on a multiple draws from a derived joint distribution of variables; that distribution is based on observed data relationships. Fully synthetic datasets create all variables this way, whereas partially synthetic datasets retain survey observations for some variables and impute other variables.

¹⁰ LEHD is a partnership between the Census Bureau and all 50 states and the District of Columbia; it produces public use data tabulations (Quarterly Workforce Indicators and an interactive web-based commuting analysis tool, OnTheMap) that are widely used by state and local governments. At its core are two AR files provided by states on a quarterly basis: (1) unemployment insurance (UI) wage records, giving the earnings of each worker at each employer; and (2) employer reports giving establishment-level data, also known as the Quarterly Census of Employment and Wages. The coverage is roughly 96 percent of private, nonfarm wage and salary employment. LEHD is in the process of integrating data on self-employed individuals and independent contractors that are not covered in the UI files but that are available from the Census Bureau Business Register, which contains the universe of all businesses, including all sole proprietorships, on an annual basis (whether the sole proprietor has employees or is a nonemployer). In addition, the LEHD project has acquired personnel records from the U.S. Office of Personnel Management so that federal workers are now also tracked in LEHD. See also Abowd, Haltiwanger, and Lane (2004).

¹¹ Partially synthetic public use datasets have already been developed for selected Census Bureau surveys (the Survey of Income and Program Participation, the Longitudinal Business Database, the LEHD dataset of AR, and ACS group quarters data) to permit release of additional microdata without jeopardizing the confidentiality of respondents. For other examples and discussions of partially synthetic data, see Abowd and Lane (2004); Abowd and Woodcock (2004, 2001); Kennickell (1997); Little (1993); Little, Liu, and Raghunathan (2004); Reiter (2005a, 2005b, 2005c, 2004a, 2004b, 2003); and Reiter and Mitra (2009).

Many AHS demographic, economic, and housing variables also appear in ACS and can therefore be used for modeling the joint distributions. Because the precise location of the unit is known, when operating within a confidential environment, one can also link census tract characteristics from the 2000 long form and ACS. Those additional variables could help in developing imputation models.

It appears likely that enough exact-matched cases are available to enable researchers to form decent imputation models. From 2000 through 2012, 16.6 million ACS interviews were completed, mostly at unique addresses.¹² In 2011, approximately 132 million housing units were in the United States; thus, ACS cases represent approximately one-eighth of all housing units (although a small fraction of ACS housing units will have been demolished by 2011). One-eighth of the 2011 AHS sample size of 180,000 units yields more than 20,000 units in the 2011 AHS that had been in the 2000–2012 ACS. Of course, not all these matches are possible, because some AHS cases are added each survey year to account for new construction, so previous AHS sample sizes were smaller (although one can use both the national AHS files and the metropolitan area AHS files), and some matching variables will be missing.

Analysts can derive models for selected AHS characteristics that are not included in ACS by examining the relationships between AHS variables that are also in ACS and those that are not. These models can then be used to create the synthetic AHS variables for ACS households—both those in the exact match universe and those only in ACS. Multiple imputations are typically done to reduce potential bias from any one draw from the joint distribution. The validity of the imputations can be tested by reference to the actual values for the exact matches, and it seems wise to focus on imputing only a few critical housing measures.

Some Data Issues

The previous section discussed some promising techniques for enhancing existing datasets for housing research. Some problems remain, however. One of these problems is undercoverage of housing units by the sampling frame that the Census Bureau uses for its household surveys. To the extent that units at the edge of habitability, or more likely units that are the result of conversion or subdivision (such as a converted garage) are missed, the statistics we use to describe our housing stock will be biased. A second type of problem is the development of new types of living quarters, such as congregate housing. As the population ages, more and more people are living in developments that cater to their needs, including housing units that do not have kitchens (because meals are provided centrally). How we should measure their growth depends on how we allow the definition of habitable housing to vary, and whether we can adequately add them to our sampling frame.

Another problem is that we measure the same concept in multiple ways and in multiple surveys. A prime example is the vacancy rate. On AHS, the definition of a vacant housing unit is quite similar to that used in the Housing Vacancy Survey.

A housing unit is vacant if no one is living in it at the time of the interview, unless its occupants are only temporarily absent. In addition, housing units where all the occupants have a usual residence elsewhere are grouped with vacant units. ... For vacant housing

¹² Duplicates can be treated as independent observations, because they will be matched to different AHS observations.

units that are not intended by their current owners for year round use (seasonal and migratory), the respondent was asked whether the construction and heating of the housing unit made it suitable for the unit to be occupied on a year-round basis. A housing unit is suitable for year-round use if it is built as a permanent structure, properly equipped, insulated, and heated as necessitated by the climate.¹³

This definition differs from the definition of a vacant housing unit used in the American Community Survey, however, because of the ACS residence rules.

The basic idea behind the ACS current residence concept is that everyone who is currently living or staying at a sample address is considered a current resident of that address, except for those staying there for only a short period of time. For the purposes of the ACS, the Census Bureau defines this short period of time as less than 2 consecutive months (often described as the 2-month rule). Under this rule, anyone who has been or will be living for 2 months or less in the sample unit when the unit is interviewed (either by mail, telephone, or personal visit) is not considered a current resident. This means that their expected length of stay is 2 months or less, not that they have been staying in the sample unit for 2 months or less. In general, people who are away from the sample unit for 2 months or less are considered to be current residents, even though they are not staying there when the interview is conducted, while people who have been or will be away for more than 2 months are considered not to be current residents. The Census Bureau classifies as vacant [a housing unit] in which no one is determined to be a current resident. (U.S. Census Bureau, 2009: chapter 6)

The implications for measuring vacancy rates derive from the ACS interview methodology. First, a questionnaire is mailed out and, starting in January 2014, a solicitation to respond is sent via the Internet as well. This questionnaire is followed about 1 month later by a telephone followup (if a phone number can be obtained), and then, for a sample of nonrespondents, about 1 more month later by a personal visit. A unit that might have been vacant at the original date of interview (month 1) may well become occupied at month 3, when the household is interviewed, yielding a lower vacancy rate than might otherwise be recorded. Thus, it is likely that vacancy rates measured by ACS differ from those measured by household surveys and from the decennial census (which uses yet a different methodology).¹⁴

Data References Appendix

American Community Survey (ACS): <http://www.census.gov/acs/www/>; see also U.S. Census Bureau, American FactFinder.

American Housing Survey (AHS): <http://www.census.gov/housing/ahs/> and <http://www.huduser.org/portal/datasets/ahs.html> (1997–2011 AHS); see also U.S. Census Bureau, American FactFinder.

¹³ Quoted from <http://www.census.gov/housing/ahs/files/>, Appendix A: 31–32.

¹⁴ See Clark (2012), Fish (2013), and Griffin et al. (2004) for comparisons with the decennial census.

Black Knight Financial Services, Inc.: <https://www.lpsdefault.com/fnds/home.asp>.

Census of Construction, 2002 and 2007 (2012 forthcoming): <http://www.census.gov/econ/census07/>;
see also U.S. Census Bureau, American FactFinder.

Construction statistics (monthly, quarterly, and annual surveys): <http://www.census.gov/econ/construction.html>.

Construction statistics (New Residential Construction, Manufactured Homes Survey, New Residential Sales, Residential Improvements, Characteristics of New Housing, Construction Price Indexes, and Value of Construction Put in Place): <http://www.census.gov/mcd/>.

CoreLogic, Inc.: <http://www.corelogic.com/solutions/property-information-analytic-solutions.aspx>.

Decennial Census of Housing, 1940–2000 (tabulations): <http://www.census.gov/housing/census/data/>.

Decennial Census of Population and Housing, selected historical decennial census population and housing counts, 1790–1990: <http://www.census.gov/population/www/censusdata/hiscendata.html>.

Decennial Census of Population and Housing, 1990: <http://www.census.gov/main/www/cen1990.html>.

Decennial Census of Population and Housing, 2000 and 2010: <http://factfinder2.census.gov>.

Fannie Mae, National Housing Survey (NHS): <http://www.fanniemae.com/portal/research-and-analysis/housing-survey.html>.

Federal Financial Institutions Examination Council, Community Reinvestment Act (CRA), Distressed and Underserved Tracts, <http://www.ffiec.gov/cra/distressed.htm>.

Federal Financial Institutions Examination Council, Home Mortgage Disclosure Act (HMDA): <http://www.ffiec.gov/hmda/default.htm>.

Federal Housing Finance Agency, government-sponsored enterprise mortgage purchases: <http://www.fhfa.gov/Default.aspx?Page=137>.

Federal Reserve Board of Governors, Flow of Funds Accounts, Net Change in Mortgages Outstanding and of Total Outstanding Mortgages: <http://www.federalreserve.gov/releases/z1/about.htm>.

Freddie Mac, House Price Index (FMHPI): <http://www.freddiemac.com/finance/fmhpi/>.

Freddie Mac, Primary Mortgage Market Survey (PMMS): <http://www.freddiemac.com/pmms/>.

Housing Vacancy Survey (HVS): <http://www.census.gov/housing/hvs/>.

Integrated Public Use Microdata Series (IPUMS): <https://www.ipums.org>.

Manufactured Homes Survey (MHS): <http://www.census.gov/construction/mhs/mhsindex.html>.

Mortgage Bankers Association, Weekly Applications Survey: <http://www.mortgagebankers.org/ResearchandForecasts/ProductsandSurveys/WeeklyApplicationSurvey>.

MPF Research: <http://www.realpage.com/apartment-market-research/rental-market-trends>.

National Association of Home Builders: <http://www.nahb.org/>.

National Association of Realtors®: <http://www.realtor.org/>.

National Council of Real Estate Investment Fiduciaries (NCREIF), Data and Products Guide: http://www.ncreif.org/public_files/NCREIF_Data_and_Products_Guide.pdf.

New Residential Construction (Building Permits Survey, Quarterly Starts and Completions by Purpose and Design, Annual Characteristics of New Housing, Length of Time from Authorization to Start and from Start to Completion, Construction Price Indexes): <http://www.census.gov/construction/nrc/>.

New York City Housing and Vacancy Survey (NYCHVS): <http://www.nyc.gov/html/hpd/html/pr/vacancy.shtml> and <http://www.census.gov/housing/nychvs/about/>.

Panel Study of Income Dynamics (PSID): <http://psidonline.isr.umich.edu/>.

Property Owners and Managers Survey (POMS): <http://www.census.gov/housing/poms/>.

Real Capital Analytics Inc.: https://www.rcanalytics.com/Public/rca_cpqi.aspx.

RealtyTrac® Inc.: <http://www.realtytrac.com/>.

Rental Housing Finance Survey (RHFS): <http://www.census.gov/hhes/rhfs/>.

Residential Energy Consumption Survey (RECS): <http://www.eia.gov/consumption/residential/>.

Residential Finance Survey (RFS): <http://www.census.gov/housing/rfs/>.

Survey of Consumer Finances (SCF): <http://www.federalreserve.gov/econresdata/scf/scfindex.htm>.

Survey of Market Absorption (SOMA): <http://www.census.gov/hhes/www/housing/soma/overview.html>.

Survey of Residential Alterations and Repairs (SORAR): <http://www.census.gov/construction/c50/c50index.html>.

United Nations Statistics Division, Compendium of Human Settlements Statistics/Compendium of Housing Statistics: <http://unstats.un.org/unsd/demographic/sconcerns/housing/housing2.htm>.

U.S. Bureau of Labor Statistics, Consumer Price Index (CPI): <http://www.bls.gov/cpi/>.

U.S. Census Bureau, American FactFinder (2000 Census of Population and Housing, 2010 Census of Population and Housing, 2005–2012 ACS, 2011 AHS, 2002–2012 Census of Construction [sector 23]): <http://factfinder2.census.gov>.

U.S. Census Bureau, Housing Unit Estimates (vintage 2012): <http://www.census.gov/popest/data/housing/totals/2012/index.html> and its methodology, at <http://www.census.gov/popest/methodology/2012-hu-meth.pdf>.

U.S. Census Bureau, Longitudinal Employer-Household Dynamics (LEHD): <http://lehd.ces.census.gov>.

U.S. Census Bureau, Research Data Centers (RDCs): <http://www.census.gov/ces/rdcresearch/index.html>.

U.S. Department of Housing and Urban Development (HUD), Homelessness Data Exchange: <http://www.hudhdx.info/>.

HUD, National Low-Income Housing Tax Credit (LIHTC) Database: <http://LIHTC.huduser.org>.

HUD, Public and Indian Housing Information Center (PIC) system and Tenant Rental Assistance Certification System (TRACS): aggregate data (A Picture of Subsidized Households) at <http://www.huduser.org/portal/datasets/assthsg.html>; a 5-percent microdata sample from those systems for researchers at <http://www.huduser.org/portal/pumnd/index.html>.

U.S. Postal Service, Vacancies: Aggregate data quarterly at the census tract level for government and not-for-profit researchers from HUD at <http://www.huduser.org/portal/datasets/usps.html>.

Zillow Home Value Index: <http://www.zillow.com/howto/api/APIOverview.htm>.

Acknowledgments

An earlier version of this article was prepared for the January 2014 conference of the Weimer School of Advanced Studies in Real Estate and Land Economics. The author thanks Arthur Cresce, George Galster, Henry Pollakowski, Mark Shroder, and Dav Vandenbroucke for their comments and suggestions.

Author

Daniel H. Weinberg is currently a visiting scholar in the Social and Decision Analytics Laboratory, Virginia Bioinformatics Institute at Virginia Polytechnic Institute and State University. He formerly held several positions at the U.S. Census Bureau, including Assistant Director for the Decennial Census and ACS (responsible for the 2010 census and the American Community Survey) and Chief of the Housing and Household Economic Statistics Division (responsible for the American Housing Survey, Housing Vacancy Survey, Residential Finance Survey, and many other housing surveys).

References

Abowd, John M., John C. Haltiwanger, and Julia I. Lane. 2004. "Integrated Longitudinal Employee-Employer Data for the United States," *American Economic Review* 94 (2): 224–229.

Abowd, John M., and Julia I. Lane. 2004. "New Approaches to Confidentiality Protection: Synthetic Data, Remote Access and Research Data Centers." In *Privacy in Statistical Databases*, edited by Josep Domingo-Ferrer and Vicenç Torra. New York: Springer-Verlag: 282–289.

Abowd, John M., and Simon D. Woodcock. 2004. "Multiply-Imputing Confidential Characteristics and File Links in Longitudinal Linked Data." In *Privacy in Statistical Databases*, edited by Josep Domingo-Ferrer and Vicenç Torra. New York: Springer-Verlag: 290–297.

———. 2001. "Disclosure Limitation in Longitudinal Linked Data." In *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*, edited by Patricia Doyle, Julia Lane, Laura Zayatz, and Jan Theeuwes. Amsterdam, The Netherlands: North-Holland: 215–277.

Andersson, Fredrik, John C. Haltiwanger, Mark J. Kutzbach, Giordano Palloni, Henry O. Pollakowski, and Daniel H. Weinberg. 2013. "Childhood Housing and Adult Earnings: A Between-Siblings Analysis of Housing Vouchers and Public Housing." U.S. Census Bureau Center for Economic Studies Discussion Paper: 13–48.

Christensen, Gunvor. 2011. "The Building and Housing Register," *Scandinavian Journal of Public Health* 39 (7): 106–108.

Clark, Sandra Lockett. 2012. "Understanding Differences in ACS and 2010 Census Information on Occupancy Status—Data Collection Methods." American Community Survey Research and Evaluation Program (May). http://www.census.gov/acs/www/Downloads/library/2012/2012_Clark_01.pdf.

Fish, Samantha. 2013. "Understanding Inconsistencies in Occupancy Status Between the 2010 Census and the 2010 American Community Survey 1-Year Estimates." American Community Survey Research and Evaluation Program (July). http://www.census.gov/acs/www/Downloads/library/2013/2013_Fish_03.pdf.

Griffin, Deborah H., Joan K. Broadwater, Theresa F. Leslie, Susan P. Love, Sally M. Obenski, and David A. Raglin. 2004. *Meeting 21st Century Demographic Data Needs—Implementing the American Community Survey*. Report 4: "Comparing General Demographic and Housing Characteristics With Census 2000." Washington, DC: U.S. Census Bureau (May). http://www.census.gov/acs/www/Downloads/library/2004/2004_Griffin_01.pdf.

Kennickell, Arthur B. 1997. "Multiple Imputation and Disclosure Protection: The Case of the 1995 Survey of Consumer Finances." Washington, DC: Federal Reserve Board of Governors. <http://www.federalreserve.gov/PUBS/oss/oss2/papers/expers95.pdf>.

Little, Roderick J.A. 1993. "Statistical Analysis of Masked Data," *Journal of Official Statistics* 9: 407–426.

Little, Roderick J.A., Fang Liu, and Trivellore E. Raghunathan. 2004. "Statistical Disclosure Techniques Based on Multiple Imputation." In *Applied Bayesian Modeling and Causal Inference From Incomplete-Data Perspectives*, edited by Andrew Gelman and Xiao-Li Meng. New York: John Wiley & Sons: 141–152.

Lubell, Jeffrey M., Mark Shroder, and Barry Steffen. 2003. "Work Participation and Length of Stay in HUD-Assisted Housing," *Cityscape* 6 (2): 207–223.

Mast, Brent D. 2012. "Comparing Public Housing and Housing Voucher Tenants With Bayesian Propensity Scores," *Cityscape* 14 (1): 55–72.

Mills, Gregory, Daniel Gubits, Larry Orr, David Long, Judie Feins, Bulbul Kaul, Michelle Wood, Amy Jones & Associates, Cloudburst Consulting, and the QED Group. 2006. *The Effects of Housing Vouchers on Welfare Families*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research; Cambridge, MA: Abt Associates Inc.

National Council of Real Estate Investment Fiduciaries (NCREIF). 2013–2014. “NCREIF Data, Index and Products Guide.” https://www.ncreif.org/public_files/NCREIF_Data_and_Products_Guide.pdf.

Olsen, Edgar O., Catherine A. Tyler, Jonathan W. King, and Paul E. Carrillo. 2005. “The Effects of Different Types of Housing Assistance on Earnings and Employment,” *Cityscape* 8 (2): 163–187.

Reiter, Jerome P. 2005a. “Estimating Identification Risks in Microdata,” *Journal of the American Statistical Association* 100: 1103–1113.

———. 2005b. “Significance Tests for Multi-Component Estimands From Multiply-Imputed, Synthetic Microdata,” *Journal of Statistical Planning and Inference* 131: 365–377.

———. 2005c. “Using CART to Generate Partially Synthetic, Public Use Microdata,” *Journal of Official Statistics* 21: 441–462.

———. 2004a. “New Approaches to Data Dissemination: A Glimpse Into the Future,” *Chance* 17 (3): 12–16.

———. 2004b. “Simultaneous Use of Multiple Imputation for Missing Data and Disclosure Limitation,” *Survey Methodology* 30: 235–242.

———. 2003. “Inference for Partially Synthetic, Public Use Microdata Sets,” *Survey Methodology* 29 (2): 181–189.

Reiter, Jerome P., and Robin Mitra. 2009. “Estimating Risks of Identification Disclosure in Partially Synthetic Data,” *The Journal of Privacy and Confidentiality* 1 (1): 99–110.

Shroder, Mark. 2002. “Does Housing Assistance Perversely Affect Self-Sufficiency? A Review Essay,” *Journal of Housing Economics* 11: 381–417.

Tatian, Peter A., and Christopher Snow. 2005. “The Effects of Housing Assistance on Income, Earnings, and Employment,” *Cityscape* 8 (2): 135–161.

U.S. Census Bureau. 2009. *American Community Survey Design and Methodology*. Report ACS-DM1 (April). Washington, DC: U.S. Census Bureau.

U.S. Department of Housing and Urban Development, Office of Policy Development and Research (HUD). 2013 (July). *HUD Research Roadmap, FY 2014–FY 2018*. http://www.huduser.org/portal/pdf/Research_Roadmap.pdf.