

Data Shop

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

The Community Advantage Program Database: Overview and Comparison With the Current Population Survey

Sarah F. Riley

HongYu Ru

Roberto G. Quercia

The University of North Carolina at Chapel Hill

Abstract

The Community Advantage Program Database (CAPD) is a unique data source that researchers can use to study the housing experiences of low-to-moderate-income and minority households in the United States. The CAPD includes data from a longitudinal panel survey of mortgage borrowers linked to monthly loan-level information and contemporaneous market valuation data and credit scores. To date, the CAPD has been used to examine a wide variety of financial and social topics for this population, including wealth and asset accumulation, mobility, postpurchase counseling, loan prepayment, neighborhood satisfaction, community involvement, and parental attitudes. After discussing the scope of the data, we compare the key demographics of the panel survey respondents with those of Current Population Survey respondents. To the best of our knowledge, no other publicly available data source combines longitudinal borrower survey data with comprehensive loan performance information.

Introduction

The Community Advantage Program Database (CAP Database, or CAPD)¹ is a unique database that combines information from the following three sources:

1. Community Advantage Panel Survey (CAP Survey, or CAPS)—an annual survey of approximately 5,200 low-to-moderate-income and minority U.S. homeowners and renters.
2. Loan-level data—origination and monthly servicing information for all original mortgage loans issued to homeowners participating in the CAP Survey.
3. Property valuation data—quarterly ZIP Code-level property evaluations for homeowners participating in the CAP Survey.

In the following text, we describe each component of the database. We conclude by examining the extent to which research based on the CAP Database can be generalized to a larger population of interest by comparing the key demographic variables of CAP Survey respondents with those of the May 2003 Current Population Survey (CPS) respondents.

Community Advantage Panel Survey

The CAP Survey fills an important research need because few existing data sources provide information about the housing experiences of the low-to-moderate-income and minority U.S. population. Moreover, the other available data generally have not been collected with this specific purpose in mind (Dietz and Haurin, 2003; Herbert and Belsky, 2008, 2006). In this section, we discuss the context and scope of the survey and data.

Background

The Community Advantage Program (CAP), from which the CAP Survey participants were drawn, was established as a partnership among the Ford Foundation, Fannie Mae, and Self-Help, a large community development financial institution located in Durham, North Carolina. Under CAP, Self-Help purchases primarily 30-year, fixed-rate mortgages (FRMs) originated through the Community Reinvestment Act-related lending activities of participating lenders. Almost all the loans made to survey participants are structured as 30-year FRMs and all were originated to finance home purchases. The CAP Survey provides an in-depth examination of the experiences of a representative sample of CAP borrowers and a similar cohort of renters.

Because all the original loans made to CAP Survey homeowners originated between 1999 and 2003, the survey spans the housing market boom that peaked in early 2006 and the bust period of subsequent declines in property values and the economic recession of 2008.

¹ The Ford Foundation funds the CAPD.

Therefore, the survey provides researchers with an opportunity to examine the experiences of low-to-moderate-income and minority borrowers under varied and changing economic conditions. It also gives them an interesting counterfactual regarding subprime lending for this population.

Scope

The CAP Survey is designed to enable researchers to evaluate the financial and social effects of homeownership among low-to-moderate-income and minority households in the United States. The homeowners in the CAP Survey constitute the primary sample of interest; the renters, who were matched to the homeowners based on geographic location and an income ceiling, form a comparison group. In addition to collecting routine demographic information, the survey collects data in the following primary areas.

Housing Experiences and Mortgage Finance

Respondents provide details about the housing unit in which they live, any improvements they have made, and their overall reasons for being satisfied or dissatisfied with their housing. They also discuss new home purchases, refinancing terms, second mortgages, home equity lines of credit, and broker involvement in the mortgage process. Both homeowners and renters report on their payment habits, including delinquencies and foreclosures, and other reasons why they may have moved. Renters in the survey also discuss their plans and expectations regarding future homeownership.

Household Wealth, Assets, and Debts

The survey collects information about all household assets and debts, including wage and nonwage income sources, retirement accounts, cash equivalents, nonliquid investments, educational and medical debt, and credit card balances. Respondents also detail the costs and causes of any bankruptcies they have experienced.

Financial Literacy, Counseling, and Savings Behavior

Researchers can examine the relationships between homeownership and self-reported financial knowledge and thriftiness as well as the effects of prepurchase and postpurchase delinquency counseling. Respondents compare their economic situations with those of their parents and discuss their money-management habits, such as credit card repayment patterns, the use of payday lenders, and whether they play the lottery or send money to friends and family.

Social and Family Outcomes and Behaviors

Respondents rate their neighborhoods and discuss their social networks, neighborhood involvement, and volunteering and voting habits. They also report on their children's academic performance, their children's behaviors, their involvement as parents and expectations of their children, and their stress levels across a variety of life dimensions.

Geocodes

The addresses of survey respondents in each year have been recorded and geocoded down to the block group level. Therefore, the survey data can easily be matched up with external databases, such as data from the 2000 census.²

Loan-Level Data

The survey data for CAP Survey homeowners can be linked to loan-level information that Self-Help has collected from the originating lenders. These data include loan characteristics, such as the amortization term and interest rate and the original loan-to-value ratio, and property characteristics, such as whether the borrower purchased a single-family home or a condominium. Available borrower characteristics include the debt-to-income ratio and both origination and updated credit scores for loans that have remained active over time. Moreover, the data include transactional histories of monthly payments, which indicate delinquencies and cures, as well as transactions for stages in the foreclosure process and changes in the company servicing the loan.

Property Valuation Data

The survey and loan-level data for CAP Survey homeowners can be linked to quarterly ZIP Code-level property valuations from 2003 to the present. These data are made available by Fannie Mae and are estimated using an internal pricing methodology that takes into consideration not only local appreciation trends but also individual property characteristics and tax appraisals.

A Comparison of the Community Advantage Panel Survey and the Current Population Survey

The ability to generalize inferences made using the CAP Database is a crucial measure of the database's potential usefulness.³ In the following text, we compare the weighted⁴ demographics of CAP Survey homeowners and renters who completed baseline interviews with those of homeowners and renters who completed the May 2003 CPS.⁵ (See exhibit 1.)

² In addition, the Census Bureau has provided industry and occupation codes for the respondents.

³ A list of selected studies that have been prepared using CAPD data is presented in the Additional Reading section. Additional technical data issues, such as potential sample selection bias due to income, racial, and gender differences between homeowners and renters, are addressed in these peer-reviewed studies.

⁴ For details about the CAPS sampling and weighting methodology and respondent attrition, see Riley and Ru (2009a, 2009b).

⁵ This administration of the CPS falls roughly at the median of the baseline homeowner interview dates. The baseline homeowner interviews occurred mainly in 2003; the matched renters, however, were selected after the baseline homeowner interviews occurred and were interviewed for the first time in 2004.

Exhibit 1

Comparison of Weighted CAP Survey and CPS Homeowner and Renter Demographics (1 of 2)

Variable	Homeowners		Renters	
	CAP Survey (N=29K) (%)	CPS (N=19M) (%)	CAP Survey (N=19K) (%)	CPS (N=19M) (%)
Gender				
Male	55.9	49.6	32.7	42.4
Female	44.1	50.4	67.3	57.6
Age				
18–25 years old	18.2	4.7	18.7	17.2
26–30 years old	22.6	6.9	14.4	15.9
31–35 years old	17.8	11.1	12.8	15.9
36–40 years old	14.1	12.7	12.1	14.1
41 years old or more	27.3	64.6	42.0	37.0
Race				
White	57.2	58.7	37.8	44.2
Black	17.8	16.8	35.6	24.6
Hispanic	21.4	16.3	22.8	23.6
Other	3.6	8.2	3.8	7.6
Educational attainment				
11th grade or less	9.7	17.5	19.9	23.4
High school graduate/GED	21.9	35.4	28.1	35.0
Some college or associate's degree	43.5	27.9	32.1	26.6
Bachelor's degree or higher	24.9	19.1	19.8	15.0
Marital status				
Married	53.3	51.3	41.3	28.0
Widowed	1.9	5.8	3.7	3.0
Divorced	17.2	20.1	19.3	20.8
Separated	2.2	5.4	6.4	10.0
Never married	25.5	17.2	29.3	38.2
Household size				
One person	21.0	25.1	31.4	36.2
Two people	27.9	25.9	26.5	18.7
Three people	19.9	17.4	17.4	17.7
Four people	16.3	17.0	13.5	14.4
Five people or more	14.8	14.7	11.2	13.0
Employment status				
Working	91.5	70.2	64.5	68.8
Looking for work (unemployed)	3.4	4.6	13.5	7.3
Retired	1.9	8.3	4.2	2.3
Out of labor force	3.2	16.8	17.7	21.6
Geographic coverage				
Midwest	24.2	21.3	10.2	19.7
Northeast	2.7	15.7	0.0	18.8
South	56.1	40.2	72.4	34.9
West	17.0	22.8	17.4	26.6

Exhibit 1

Comparison of Weighted CAP Survey and CPS Homeowner and Renter Demographics (2 of 2)

Variable	Homeowners		Renters	
	CAP Survey (N=29K) (%)	CPS (N=19M) (%)	CAP Survey (N=19K) (%)	CPS (N=19M) (%)
Income				
Less than \$20,000	10.8	23.8	47.2	49.9
\$20,000–\$24,999	13.9	11.9	11.8	12.1
\$25,000–\$29,999	15.8	12.8	15.3	11.9
\$30,000–\$39,999	25.2	26.5	14.0	17.3
\$40,000–\$49,999	21.2	10.9	7.4	6.2
\$50,000 or more	13.1	14.1	4.4	5.5

CAP = Community Advantage Panel.

CPS = Current Population Survey.

GED = general equivalency diploma.

Notes: All differences are statistically significant with $p < 0.05$. The sample size indicated for each group is the sum of the survey weights.

Data Preparation

The CPS is a survey of 50,000 households that the Census Bureau and Bureau of Labor Statistics conduct monthly. The survey is designed to represent the noninstitutionalized civilian population in the United States. Thus, it collects information about demographics and other household characteristics to provide an integrated picture of the U.S. labor force and its experiences. Because the CAP Survey represents households meeting Self-Help lending criteria and certain age requirements, we subset the CPS using similar criteria so that respondents of the two surveys can be more readily compared.

Specifically, because the survey respondent for the CAP Survey is the head of the household, we restrict the CPS to the reference person of each household, whom we identify as the person whose name is on the property deed or rental contract for that household. In addition, we exclude full-time students and respondents over the age of 65 because efforts were made to exclude such individuals from the CAP Survey pool.

We further subset the CPS to include only those households that meet a simplified version of the CAP lending criteria. That is, the CPS household income must but be no greater than 80 percent of the Area Median Income (AMI) at the metropolitan statistical area level or no greater than 115 percent of the AMI if the respondent is a minority.

Homeowners

The first two columns of exhibit 1 provide demographic proportions for the CAP Survey homeowners who completed the baseline interview and CPS homeowners. Compared with the CPS homeowners, the CAP Survey homeowners are more likely to be male (56 versus 50

percent) and more likely to be Hispanic (21 versus 16 percent). The proportions of White and Black respondents in these two samples are roughly comparable, however.

The demographic data also show that CAP Survey homeowners tend to be younger and more educated than CPS homeowners and are also more likely to be employed. About 40 percent of CAP Survey homeowners were 30 years old or younger when interviewed, compared with only 12 percent of CPS homeowners. In addition, only 27 percent of CAP Survey homeowners were 41 years old or older when interviewed, while 65 percent of CPS homeowners were in that age cohort when surveyed. Regarding educational attainment, CPS homeowners are nearly twice as likely as CAP Survey homeowners not to have finished high school. Moreover, CPS homeowners are only 70 percent as likely as CAP Survey homeowners to have completed at least some college. More than 90 percent of CAP Survey homeowners were employed as of the baseline interview, compared with 70 percent of CPS homeowners.

Despite these differences, CAP Survey homeowners and CPS homeowners are very similar regarding their overall household size distribution, and both sets of respondents have about a 50-percent likelihood of being married. The income distribution of CAP Survey homeowners is also similar to that of CPS homeowners, except that only 11 percent of CAP Survey homeowners are in the lowest annual income bracket (less than \$20,000), compared with 24 percent of CPS homeowners.

The greatest difference between CAP Survey homeowners and CPS homeowners concerns geographic coverage. The CAP Survey has very little coverage in the Northeast and overrepresents the South.

Renters

The second set of columns in exhibit 1 provides demographic proportions for the CAP Survey renters who completed baseline interviews and renters in the CPS. Compared with the CPS renters, the CAP Survey renters are more likely to be female (67 versus 58 percent), less likely to be White (38 versus 44 percent), and more likely to be Black (36 versus 25 percent). Despite these differences, the proportions of Hispanic respondents in these two samples are roughly comparable.

The data also show that CAP Survey renters tend to be older and more educated than CPS renters. About 42 percent of CAP Survey renters were 41 years old or older when interviewed, compared with 37 percent of CPS renters. The CAP Survey renters are 4 percentage points less likely to not have completed high school. Moreover, 52 percent of CAP Survey renters had completed at least some college, compared with 42 percent of CPS renters.

CAP Survey renters were slightly less likely to be employed than CPS renters were (65 versus 69 percent) but nearly twice as likely to be unemployed (that is, not working but looking for work) than CPS renters were. It appears that the CAP Survey renters who were not working were more likely to be looking for work, while comparable CPS renters were more likely instead to have opted out of the labor force.

The income and household size distributions are similar for CAP Survey renters and CPS renters, although the former are more likely to have a second person living in the household in addition to the survey respondent. The CAP Survey renters are also more likely than CPS renters to be married (41 versus 28 percent) and less likely to have separated or never been married than their CPS counterparts.

As with the homeowners, the greatest difference between CAP Survey renters and CPS renters concerns geographic coverage. More than 70 percent of CAP Survey renters are located in the South, compared with 35 percent of CPS renters. Moreover, 19 percent of CPS renters live in the Northeast but none of the CAP Survey renters do.

Conclusion

The CAP Database is a unique resource for researchers interested in the housing experiences of low-to-moderate-income and minority households. The survey data include a broad range of social and financial outcomes collected during an unusual period of U.S. economic history and are linked with rich proprietary loan-level data concerning mortgage performance and housing appreciation. The CAP Survey participants are similar to comparable CPS respondents regarding household size, income distribution, and minority representation, although, compared with CPS respondents, CAP Survey participants tend to be slightly more educated, to demonstrate greater attachment to the workforce, and to be much more likely to live in the South.

In the years since the study began, some respondents have moved from their original locations and/or changed tenure status, thereby blurring the distinction between the original groups of renters and homeowners. Because the survey continues to follow respondents after a move and/or a tenure change, respondent mobility adds additional richness by expanding the range of topics that can be examined using the data. As is common for surveys, sample sizes may also be a statistical concern in some cases. Survey modules that intentionally target a subset of respondents, such as mobility or parenting questions, may provide data for several hundred respondents rather than for several thousand.

A version of the CAP Database that omits the personally identifying information of respondents is currently being prepared for public use. In all likelihood, researchers will be able to access most of these de-identified data. Confidential and proprietary information may be made available through special arrangements. Researchers who are interested in using the data or who want to receive more information about the upcoming public-use data sets should contact the Center for Community Capital at The University of North Carolina at Chapel Hill.

Acknowledgments

The authors thank Mark Lindblad and Kim Manturuk at the Center for Community Capital for their helpful comments.

Authors

Sarah F. Riley is a data analyst at the Center for Community Capital, The University of North Carolina at Chapel Hill.

HongYu Ru is a research assistant at the Center for Community Capital, The University of North Carolina at Chapel Hill.

Roberto G. Quercia is the director of the Center for Community Capital and a professor in the Department of City and Regional Planning, The University of North Carolina at Chapel Hill.

References

- Dietz, Robert D., and Donald R. Haurin. 2003. "The Social and Private Micro-Level Consequences of Unemployment," *Journal of Urban Economics* 54: 401–450.
- Herbert, Christopher E., and Eric S. Belsky. 2008. "The Homeownership Experience of Low-Income and Minority Households: A Review and Synthesis of the Literature," *Cityscape* 10 (2): 5–60.
- . 2006. *The Homeownership Experience of Low-Income and Minority Families: A Review and Synthesis of the Literature*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.
- Riley, Sarah, and HongYu Ru. 2009a. *Community Advantage Panel Survey Technical Sampling Report: Owners, 2003–2008*. Chapel Hill, NC: University of North Carolina, Center for Community Capital.
- . 2009b. *Community Advantage Panel Survey Technical Sampling Report: Renters, 2004–2008*. Chapel Hill, NC: University of North Carolina, Center for Community Capital.

Additional Reading⁶

- Ding, Lei, Roberto G. Quercia, and Janneke Ratcliffe. 2008. "Post-Purchase Counseling and Default Resolutions Among Low- and Moderate-Income Borrowers," *Journal of Real Estate Research* 30 (3): 315–344.
- Grinstein-Weiss, Michal, Johanna K.P. Greeson, Yeong H. Yeo, Susanna S. Birdsong, Mathieu R. Despard, and Roberto G. Quercia. 2009. "The Impact of Low- and Moderate-Wealth Homeownership on Parental Attitudes and Behavior: Evidence From the Community Advantage Panel," *Children and Youth Services Review* 31: 23–31.

⁶ For a list of working papers currently using the CAPD, see the Center for Community Capital website: <http://www.ccc.unc.edu/research.php>.

Rohe, William M., Roberto Quercia, and Shannon Van Zandt. 2007. "The Social-Psychological Effects of Affordable Homeownership." In *Chasing the American Dream: New Perspectives on Affordable Homeownership*, edited by William M. Rohe and Harry Watson. Ithaca, NY: Cornell University Press.

Spader, Jonathan S., and Roberto G. Quercia. 2008. "Does Homeownership Counseling Affect the Prepayment and Default Behavior of Affordable Mortgage Borrowers?" *Journal of Policy Analysis and Management* 27 (2): 304–325.

Spader, Jonathan S., and Roberto G. Quercia. 2008. "Mobility and Exit From Homeownership: Implications for Community Reinvestment Lending," *Housing Policy Debate* 19 (4): 675–709.

Stegman, Michael A., Roberto G. Quercia, and Walter R. Davis. 2007. "The Determinants of Home Price Appreciation Among Community Reinvestment Homeowners," *Housing Studies* 22 (3): 381–408.