

How Well Does the Current Population Survey Represent California?

Deborah Reed
reed@ppic.org

Public Policy Institute of California

March 2001

Acknowledgements: This project has benefited greatly from the skilled programming assistance of Jennifer Cheng. Hans Johnson, Richard Lovelady, and Michael Teitz provided helpful comments on an earlier draft. Gregory Weyland of the Census Bureau answered numerous questions on CPS design and methods.

Summary

The Annual Demographic File (March file) of the Current Population Survey (CPS) is perhaps the most important data source for annual information on social and economic trends in the state of California. Yet the CPS was designed to produce national estimates rather than state-level estimates. This project addresses two main issues in the March CPS from 1970 onward. First, has the California subsample of the CPS been representative of the state? Second, can California weights improve the representation in the sample?

The study finds that the CPS sample has represented fairly accurately the characteristics of the state population in terms of sex, age, and race/Hispanic origin. However, there is clearly room for improvement. For example, in the 1998 survey, the CPS reported too many Native Americans, Asians, and youth living in the state. However, the regional distribution of the population within California and several social indicators (poverty, household income, and education) were accurately represented.

To improve population representation, California weights were created based on independent estimates of the state population by sex, age, and race/Hispanic origin. By construction, the California weights improved the sample distribution of basic demographic characteristics. However, the regional distribution of the population within California and the social indicators were not substantially changed by the California weights.

The most important result of this study is that for most years several socioeconomic indicators appear to be substantially unaffected by using the California weights. The results suggest the validity of state-level estimates that use the official national weights. However, the California weights appear to moderately reduce the impact of survey redesigns in 1985 and 1994 on the estimates of trends in poverty and family income. Researchers interested in using the California weights are referred to the last chapter of this report for a description of the weights data sets created for this study and how to access them from the Institute.

Contents

Summary	i
1. INTRODUCTION	1
2. HOW WELL DOES THE CURRENT POPULATION SURVEY REPRESENT CALIFORNIA?	3
Sex, Age, and Race/Hispanic Origin	3
Regional Distribution	5
Social Indicators	6
3. CAN REPRESENTATION BE IMPROVED WITH CALIFORNIA WEIGHTS?	9
Sex, Age, and Race/Hispanic Origin	9
Regional Distribution	13
Social Indicators	13
4. CONSTRUCTION OF CALIFORNIA WEIGHTS	17
5. ACCESSING, USING, AND UPDATING THE CALIFORNIA WEIGHTS	24
When to Use Weights and Which Weights to Use	24
How to Access California Weights from PPIC	25
Future Updates	26
Bibliography	27

1. Introduction

The Current Population Survey (CPS) is a monthly survey of civilian noninstitutional households in the United States collected by the U.S. Bureau of the Census. The purpose of the Annual Demographic File – a special supplement to the CPS surveyed annually in March – is to study social and economic trends for individuals, households, and families, including living arrangements, fertility, marriage, education, earnings, and income. The March survey is publicly available every year, starting in 1968. Currently, the survey has about 50,000 households nationally and about 5,000 households in California.

The CPS is the major source for annual data on social and economic indicators in California. For example, official measures of state poverty rates rely on data from the March file of the CPS. Similarly, most California studies of trends in earnings, family income, income distribution, household composition, marriage, and fertility use data from the CPS.¹

The California subsample of the CPS may not accurately represent the state. The survey was not designed to be representative at the state level, but rather at the national level.² To improve national representation, the Census Bureau creates weights so that the sex, age, race, and Hispanic origin characteristics of the sample will match independent national population estimates. Each observation is assigned a weight. When the weights are summed for observations belonging to a group defined by sex, age, race and/or Hispanic origin, the sum should be roughly equivalent to independent estimates of the national population for that group. Since the weights are created at the national level, if a demographic group is underrepresented in the national sample, members of the group living in California will be assigned high weights, even if that group is not underrepresented in the state sample. For example, Native Americans were underrepresented in the 1998 CPS, and thus the national weights are relatively high for this group. Applying the weights to the California subsample leads to an overestimate of the number of Native Americans in California compared to independent estimates of the state population.

Furthermore, the independent population estimates have a fair degree of estimation error, but the weights from prior years are not reevaluated following a decennial Census when more accurate population information is available. Finally, while the CPS weighting process has been improved over the past thirty years, the weights from earlier years have not been recalculated to reflect new methods. For example, estimates of undocumented immigration were not included in the weights calculations until 1986, and estimates of decennial Census undercounts from the Post Enumeration Survey were not included until 1994.

This project has two main goals. The first is to compare the California subsample of the March file of the CPS to independent estimates of the characteristics of the state

¹ The Census Bureau plans to fully implement the American Community Survey (ACS) in 2003. If implemented, the ACS will replace the CPS as the major source of annual social and economic data for California. See www.census.gov/acs/www for a description of this valuable survey.

² The random sampling method for the March CPS has been designed at the state level since 1985. Prior to 1985, sampling was at the regional level, and California was sampled with the other states in the Western region. Beginning in 1978, the weighting algorithm included an adjustment for total population in each state.

population to determine how accurately the survey has represented California from 1970 onward. The second goal is to produce weights for the California subsample of the CPS, using independent state population estimates by sex, age, and race/Hispanic origin from 1970 through 1998. The California weights will reflect the most recent weighting methodology used for the CPS and the most comprehensive state population estimates.

This study does not provide a comprehensive manual for using the CPS to study California. Researchers should consult the annual Census Bureau technical documentation before using the CPS data. In particular, users should note that the CPS design and methods changed several times over the last three decades.³ For example, the most recent major redesign of the March survey took effect in 1994. Poverty and income statistics from earlier surveys are not directly comparable with those from 1994 and later. Another concern is the small number of California observations, leading to low precision of state estimates, especially for subpopulations (e.g., Asians in California). Using a three-year moving average can substantially reduce the size of estimated confidence intervals.

The next chapter of this document compares the CPS sample to independent estimates of the population. The third chapter makes the same comparisons using California weights. The fourth chapter describes the method used to produce the California weights. The final chapter discusses the proper use of weights in the CPS, explains how to access the California weights from the Public Policy Institute of California, and describes our plans for an update of this study following the release of the 2000 decennial Census microdata, currently scheduled for 2003.

³ Reed, Glenn Haber, and Mameesh (1996), Appendix A, discuss design changes in the CPS that are relevant to the measurement of income trends.

2. How Well Does the Current Population Survey Represent California?

This chapter presents comparisons between the California sample distributions in the CPS and independent estimates of population characteristics. The main conclusion from these comparisons is that the CPS provides a fairly accurate description of the California population.

Sex, Age, and Race/Hispanic Origin

The California Department of Finance provides estimates of the California population by sex, age, and race/Hispanic origin.⁴ I adjusted these estimates to reflect the civilian noninstitutional population sampled in the CPS. The adjustment is based on national Armed Forces and institutionalized rates by sex, age, and race/Hispanic origin from the decennial Census (see Chapter 4 for details). After adjusting, the Department of Finance reported roughly 33,116,000 civilian noninstitutional persons living in the state in 1998 (see Table 1, top of column 1).

Compared to the Department of Finance estimates, the sample in the CPS is slightly too small. The CPS has two main weights: the final weight and the March supplement weight. The final weight is calculated at the individual level while the March weight adjusts for consistency among family members.⁵ Using the final weight,⁶ the total sample is missing about 400,000 people or just over 1 percent (see top of column 2). Using the March supplement weight,⁷ the total sample is missing about 209,000 people, just over half a percent (see top of column 3).

The CPS sample population is within a few percentage points of the Department of Finance estimates for many sex/race/Hispanic groups (see Table 1, second panel). The CPS does overestimate the number of Native Americans by 20 percent for males and by 14 to 16 percent for females as well as the number of Asians by 4 to 5 percent for males and by 8 to 9 percent for females. The CPS underestimates the number of black females by 6 percent.

The CPS sample population tends to be younger than suggested by the Department of Finance estimates (see Table 2.1, third panel). The CPS overestimates the population aged 10 to 18 years by 5 percent and the population aged 19 to 26 by 6 to 8 percent. The CPS underestimates the population aged 47 to 56 as well as the population aged 57 to 66 years by 7 percent.

⁴ See the Department of Finance website at <http://www.dof.ca.gov/html/Demograp/race.htm>. According to Mary Heim at the Department of Finance, for the years 1970-1996 the population counts were estimated, and for the years 1997 and beyond they were projections (as of October 1999). The counts reflect the population on July 1 of each year.

⁵ See Chapter 4 for a further description of these weights and Chapter 5 for a discussion of the use of these weights.

⁶ The final weight is sometimes referred to as the basic weight. From 1976-1988, the CPS code is B-WEIGHT. In later years the code is A-FNLWGT. The Unicon Corporation code is WGTFNL.

⁷ The March supplement weight is coded in the CPS as WEIGHT before 1976 and as MARSUPWT in 1976 and later years. The Unicon Corporation code is WGT.

**Table 2.1. Independent and CPS Population Counts for California for 1998
by Sex, Age, and Race/Hispanic Origin, National Weights**

	Dept. of Finance ^a (1000s)	Final weight (% diff)	March weight (% diff)	Census Bureau ^a (1000s)	Final weight (% diff)	March weight (% diff)
ALL	33,116	-1	-1	32,280	2	1
SEX, RACE/HISPANIC ^b						
White, male	8,425	-3	-2	8,078	1	2
Hispanic, male	5,110	0	1	5,104	0	1
Asian, male	1,819	4	5	1,754	8	9
Black, male	1,079	0	1	1,019	6	7
Native American, male	96	20	20	91	27	27
White, female	8,635	-3	-3	8,238	1	1
Hispanic, female	4,815	-1	0	4,910	-3	-2
Asian, female	1,880	8	9	1,898	7	8
Black, female	1,155	-6	-6	1,094	-1	0
Native American, female	101	16	14	95	23	21
AGE						
0-9 years	5,660	-6	-4	5,261	2	3
10-18 years	4,204	5	5	4,100	8	8
19-26 years	3,373	6	8	3,539	1	3
27-36 years	5,329	-2	-2	5,326	-1	-2
37-46 years	5,341	3	3	5,193	6	6
47-56 years	3,785	-7	-7	3,586	-2	-2
57-66 years	2,333	-7	-7	2,165	0	1
67 and older	3,090	-4	-4	3,110	-5	-5

Source: Author's calculations based on data from the California Department of Finance, the U.S. Census Bureau, the 1980 and 1990 decennial Censuses, and the 1998 March CPS.

^a Adjusted to remove Armed Forces and institutionalized populations not represented in the CPS (see Chapter 4 for details).

^b Hispanics of any race are included in the count of Hispanics and are not counted in the other race categories.

The U.S. Census Bureau also provides estimates of the state population by sex, age, race, and Hispanic origin.⁸ Since 1990, the main difference between the Census Bureau and the Department of Finance calculations has been the estimate of domestic migration. The Census Bureau matches tax returns across years to estimate migration between states. The Department of Finance uses driver's license address changes to estimate interstate migration. The Department of Finance believes that the Census Bureau estimates are too low because tax returns are often filed many months after a move. Also, the number of tax returns in California that cannot be matched and are excluded as "first time filers" is too large according to the Department of Finance.⁹ Recent data from the 2000 decennial Census suggest that the true population count lies between that of the Department of Finance and that of the Census Bureau. It is beyond the scope of this report to determine which estimates are better. Rather, I have provided a parallel analysis with both population estimates.

After adjusting to the civilian noninstitutional population, the Census Bureau estimates have 836,000 fewer Californians than do the Department of Finance estimates. Compared to the Census Bureau estimates, the California sample of the CPS is 1 to 2 percent too large. The difference is probably explained by undercount estimates. Beginning in 1994, the CPS weights were adjusted to reflect "undercount" in the national population estimates. However, the Census Bureau population estimates reported in Table 1 do not adjust for undercount.¹⁰

Similar to the Department of Finance estimates, the Census Bureau population estimates suggest that the CPS sample overestimates the number of Native American males by 27 percent and the number of Native American females by 21 to 23 percent. The number of Asian males is 8 to 9 percent too high and the number of Asian females is 7 to 8 percent too high. The CPS sample also overestimates black males by 6 to 7 percent.

The age distribution in the CPS is closer to the estimates of the Census Bureau than those of the Department of Finance. However, there is still a tendency for the CPS to over-represent the young. Compared to the Census Bureau estimates, the CPS sample has 8 percent too many people aged 10 to 18 and 5 percent too few people aged 67 and older.

The comparisons in Table 2.1 are based on the main categories that the Census Bureau uses to weight the CPS sample. The table shows that while the CPS is broadly representative of California, there appears to be room for improvement.

Regional Distribution

It is also possible to use Department of Finance¹¹ and Census Bureau¹² estimates to investigate the regional distribution of the state population.

⁸ See the U.S. Census Bureau website at <http://www.census.gov/population/www/estimates/statepop.html>. The reference date for these estimates is July 1 of each year.

⁹ This information was provided by Mary Heim at the Department of Finance.

¹⁰ The Department of Finance estimates also do not adjust for the undercount. See Chapter 4 for a further discussion of undercount adjustments.

¹¹ See <http://www.dof.ca.gov/html/Demograp/e-2.xls>.

¹² See http://www.census.gov/population/estimates/county/co-98-1/98C1_06.txt.

According to the Department of Finance, 16,256,000 people lived in the Los Angeles area in 1998. I adjusted the regional estimates by removing the percent Armed Forces and/or institutionalized population in the region calculated from the 1990 decennial Census, and this left 16,003,000 civilian noninstitutional people in the Los Angeles area (see Table 2.2, row 1). The CPS sample is fairly close to the Department of Finance estimate – only 1 percent too small using the final weight and 1 percent too large using the March weight. The CPS sample for the San Francisco area is the same as the Department of Finance estimate. Using the final weight, the CPS underestimates the population in the San Diego area by 4 percent, but the number is correct using the March weight. The population in the Sacramento region is overestimated by 2 to 3 percent relative to the Department of Finance estimates.

Comparisons with the Census Bureau regional population estimates show that the CPS sample does not match quite as well. Using the final weights and the March weights, each of the regions is too large by 2 to 4 percent in the CPS, except for San Diego, which is 3 percent too small using the final weights.

**Table 2.2. Independent and CPS Population Counts in 1998
by Region within California, National Weights**

	Dept. of Finance ^a (1000s)	Final weight (% diff)	March weight (% diff)	Census Bureau ^a (1000s)	Final weight (% diff)	March weight (% diff)
Los Angeles area	16,003	-1	1	15,535	2	4
San Francisco area	6,628	0	0	6,453	2	2
San Diego area	2,676	-4	0	2,631	-3	2
Sacramento area	1,663	3	2	1,654	4	3

Source: Author's calculations based on data from the California Department of Finance, the U.S. Census Bureau, the 1990 decennial Census, and the 1998 March CPS.

^a Adjusted to reflect civilian noninstitutional population by calculating the percent Armed Forces and/or institutionalized from the 1990 decennial Census for each geographic region and applying that percent to the 1998 population estimate.

Notes: The Los Angeles area includes the counties of Los Angeles, Orange, Riverside, San Bernardino, and Ventura. The San Francisco area includes the counties of Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano, and Sonoma. The San Diego area is San Diego County. The Sacramento area includes Sacramento, Placer, El Dorado, and Yolo counties.

Social Indicators

For 1990, I compared the CPS sample distribution of a variety of social indicators with that of the decennial Census survey using the Census Public Use Microdata Sample (PUMS). Even at the national level, the CPS and Census samples provide different estimates. The 1990 Census was conducted primarily by mail, and respondents were asked about 8 specific types of income. The 1990 CPS was conducted primarily by telephone using trained survey takers, and respondents were asked about more than 20 specific types of income. For these

reasons, one might expect the CPS to have more accurate poverty and income measures than the Census.

The first columns of Table 2.3 compare measurements of the poverty rate, household income, and education in the Census and CPS at the national level. The U.S. poverty rate as measured by the CPS is slightly lower than that measured by the Census. This may reflect either better measurement of income in the CPS or a more inclusive sample in the Census. The distribution of household income in the two surveys matches almost perfectly.¹³

Table 2.3. Social Indicators in the 1990 Census and CPS, National Weights

	United States			California		
	Census (%)	Final weight (%)	March weight (%)	Census (%)	Final weight (%)	March weight (%)
POVERTY RATE ^a	13.1	12.8	12.9	12.6	12.9	12.9
HOUSHOLD INCOME ^a						
Less than \$20,000	22	23	23	20	22	22
\$20,000 - \$39,999	30	30	30	26	27	27
\$40,000 - \$59,999	23	23	23	22	22	23
\$60,000 - \$99,999	18	18	18	21	20	20
\$100,000 and over	7	7	7	11	9	9
EDUCATION ^b						
Less than H.S. diploma	16	14	14	20	18	18
H.S. diploma	30	37	37	20	27	28
Some college	30	24	24	34	27	27
Bachelor's	16	16	16	18	17	17
Advanced degree	8	9	9	9	11	11
Twelve years (adjusted)	33	37	37	24	27	28

Source: Author's calculations based on data from the 1990 decennial Census and the 1990 March CPS.

^a The 1990 surveys ask about income in 1989, so the poverty and income measures reflect 1989 levels.

^b Educational attainment reported for persons ages 25 to 50. The education distribution in the Census differs from that of the CPS in part because of a difference in the questions asked. In 1990, the Census asked about completion of degrees, whereas the CPS asked about years of schooling. In the final row, "Twelve years" adjusts Census data to include persons finishing twelve years of school as well as those with a high school diploma.

Notes: Percentages may not add to 100 due to rounding.

¹³ Household income is weighted by the number of persons in the household (by using the sum of person weights). That is, 7 percent of people lived in households with \$100,000 or more income. I adjusted for the number of people in the household (n) by multiplying household income by $2/\sqrt{n}$.

The distribution of education is quite different in the two surveys. In part, this reflects differences in the survey questions. The CPS asked whether the respondent had completed twelve years of schooling. The Census asked whether the respondent obtained a high school diploma (or equivalent). The final row of the table shows the percent reporting "twelve years" when the Census calculation is adjusted to also include adults with twelve years of schooling but no high school diploma. By this measure, the Census and CPS are more similar; but at the national level, the CPS shows about 4 percent more people in this category than the Census.

At the state level, the Census is preferred to the CPS because the sampling procedure is representative of the state and the sample is quite large. In 1990, the California subsample of the Census had over 1 million observations. The CPS sample had just over 14,000 observations.

The final three columns of Table 2.3 compare the Census measurements to those of the CPS at the state level. The California poverty rate as measured by the two surveys is similar, but the Census rate is slightly lower. The distribution of household income is fairly close in the two surveys but not quite as close as at the national level. The difference between the two surveys in measuring the distribution of education shows the same pattern as found at the national level. Similar to the national statistics, the share of people with twelve years of schooling is about 4 percentage points higher in the CPS than in the Census (last row of table).

Overall, the comparisons between the CPS and the independent population estimates, as well as the comparisons between the CPS and the decennial Census, suggest that the CPS sample for California is representative of the state to a large degree. However, there is clearly room for improvement. In particular, the CPS sample over-represents Native Americans, Asians, and youth but represents fairly accurately the regional distribution of the population and the distribution of several social indicators.

3. Can Representation be Improved with California Weights?

One strategy for improving the representation of the California subsample of the CPS is to create sample weights based on independent population counts for California. I used the population estimates from the California Department of Finance and the United States Census Bureau to create two sets of sample weights by sex, age, and race/Hispanic origin. To create the weights, I followed the same basic strategy used for the official national weights in the CPS in 1998. Details on the weighting methodology are in Chapter 4. This chapter evaluates whether representation is improved with the California weights.

Sex, Age, and Race/Hispanic Origin

The first three columns of data in Table 3.1 are based on the Department of Finance population estimates. When the final and March weights are constructed from the Department of Finance estimates, the total population in the CPS sample matches that of their estimates by construction. Similarly, the weights were created to match the Department of Finance estimates of the distribution of sex, age, and race/Hispanic origin. However, the CPS sample has so few Native Americans under the age of eighteen (eleven boys and five girls in 1998) that I combined the sexes when I created the weights. Thus, the total number of Native Americans is correct, but the number of boys is slightly too high and the number of girls is slightly too low by about 8 percent or roughly 8,000 people.¹⁴

In constructing the weights, I used three-year age groupings for persons under age 25 and five-year age groupings for those age 25 and over (see Chapter 4 for details). The age categories reported in Table 3.1 overlap and intersect the age categories used in constructing the weights. Therefore, they do not provide a perfect match to the Department of Finance estimates, but the match is quite close. The California weights based on the Census Bureau population estimates (final three columns of Table 3.1) show the same pattern. The distribution of sex, and race/Hispanic origin matches exactly except for Native Americans, for whom the total number is correct although there are too many boys and too few girls. The age distribution matches closely but not perfectly.

Comparing Table 3.1 to its counterpart Table 2.1 reveals that using California weights has substantially improved representation in almost every cell. In particular, the over-representation of Native Americans, Asians, and youth has been corrected. This should come as no surprise. Because the weights were constructed based on the independent estimates by sex, age, and race/Hispanic origin, the newly weighted sample closely matches those estimates by construction.

¹⁴Because the March weight is constructed at the family level, the numbers are off by somewhat more using the March weights.

Table 3.1. Independent and CPS Population Counts in 1998 by Sex, Age, and Race/Hispanic Origin, California Weights

	<u>Dept. of Finance</u>			<u>Census Bureau</u>		
	Estimate ^a (1000s)	CA Final weight (% diff)	CA March weight (% diff)	Estimate ^a (1000s)	CA Final weight (% diff)	CA March weight (% diff)
ALL	33,116	0	0	32,280	0	0
SEX, RACE/HISPANIC ^b						
White, male	8,425	0	0	8,078	0	0
Hispanic, male	5,110	0	0	5,104	0	0
Asian, male	1,819	0	0	1,754	0	0
Black, male	1,079	0	-1	1,019	0	0
Native American, male	96	8	13	91	7	12
White, female	8,635	0	0	8,238	0	0
Hispanic, female	4,815	0	0	4,910	0	0
Asian, female	1,880	0	0	1,898	0	0
Black, female	1,155	0	0	1,094	0	0
Native American, female	101	-8	-8	95	-7	-7
AGE						
0-9 years	5,660	1	0	5,261	1	0
10-18 years	4,204	-1	0	4,100	0	0
19-26 years	3,373	1	1	3,539	0	0
27-36 years	5,329	-1	-2	5,326	-2	-2
37-46 years	5,341	1	1	5,193	1	1
47-56 years	3,785	1	1	3,586	2	1
57-66 years	2,333	-1	-1	2,165	0	0
67 and older	3,090	-1	-2	3,110	-1	-2

Source: Author's calculations based on data from the California Department of Finance, the U.S. Census Bureau, the 1980 and 1990 decennial Censuses, and the 1998 March CPS.

^a Adjusted to remove Armed Forces and institutionalized populations not represented in the CPS (see Chapter 4 for details).

^b Hispanics of any race are included in the count of Hispanics and are not counted in the other race categories.

To compare the national weights with the California weights for population cells described by sex, age, and race/Hispanic origin (i.e., the intersection of these categories), I calculated the difference between the independent population estimate and the CPS population for every demographic group. When possible, the demographic groups were defined by sex and single year of age within each of five groups: white non-Hispanic, Hispanic, Asian non-Hispanic, black non-Hispanic, and Native American non-Hispanic.¹⁵ For each year from 1970 to 1998, Table 3.2 reports the sum across all population cells of the absolute value of the difference between the CPS estimate and the independent estimate (in millions). For example, the sum of the differences in each cell between the CPS with national final weights and the Department of Finance population estimates was 6.9 million in 1998. Using the California final weight, the sum was 5.7 million. Table 3.2 shows that for every year for both final and March weights, the California weights improve on the national weights. Again, this result is not surprising, given that the California weights were constructed based on the independent population estimates. Before turning to other population measures, I should emphasize that the excellent performance of the Census Bureau-based California weights in the years before 1990 is due to the aggregation of cells in the Census Bureau population estimates. Between 1970 and 1980, the Census Bureau population estimates are not detailed by sex, race, or Hispanic origin; and age is collapsed into aggregated cells. The summation in Table 3.2 only uses 23 cells during these years, compared to 180 cells available in the 1990s. With fewer cells, the sum of differences by cell is necessarily lower, but the actual representation in the CPS is less accurate. Similarly, between 1981 and 1990, the Census Bureau population estimates have information on sex, race, and Hispanic origin; but age is collapsed into five-year age groups. The Department of Finance population estimates (and the Census Bureau estimates after 1989) provide more detail – allowing for the creation of weights that more closely match the population of California, even though the sum of absolute differences between the CPS and the population estimates is higher.

¹⁵ The cells are as described in the text for the Department of Finance estimates from 1988b-1998 and for the Census Bureau estimates from 1990-1998. Before 1988b, Asians were grouped with Native Americans in an “other” category because Asians were not separately identified in the CPS in those years. For the Department of Finance estimates in 1970, I did not use a race or Hispanic origin control because the CPS does not specify Hispanic origin in that year. For the Census Bureau population estimates from 1981 to 1989, age is measured in five-year age groups. For the Census Bureau population estimates from 1970 to 1980, there is no sex specification, age is measured in larger groupings, and there is no race or Hispanic origin specification.

Table 3.2 – Sum of Absolute Differences (millions)

	<u>Department of Finance</u>				<u>Census Bureau</u>			
	Final weight ^a	CA Final weight	March weight	CA March weight	Final weight ^a	CA Final weight	March weight	CA March weight
1970		1.8	2.3	1.8		0.0	1.0	0.0
1971		2.8	3.2	2.8		0.0	0.8	0.0
1972		2.9	3.1	2.9		0.0	0.7	0.0
1973		2.9	3.3	3.0		0.0	0.7	0.0
1974		3.0	3.4	3.0		0.0	0.7	0.0
1975		3.0	3.4	3.1		0.0	0.9	0.0
1976	3.5	3.1	3.5	3.2	0.8	0.0	0.8	0.0
1977	3.5	3.1	3.5	3.1	1.2	0.0	1.2	0.0
1978	3.7	3.3	3.7	3.3	1.3	0.0	1.3	0.0
1979	3.8	3.2	3.8	3.2	1.0	0.0	1.0	0.0
1980	3.6	3.1	3.6	3.1	1.0	0.0	0.8	0.0
1981	3.7	3.1	3.8	3.1	1.9	0.6	2.1	0.6
1982	4.3	3.4	4.1	3.4	2.4	0.8	2.2	0.8
1983	4.2	3.3	4.0	3.4	2.2	0.7	2.2	0.7
1984	4.3	3.5	4.0	3.5	2.1	0.8	2.1	0.7
1985	4.8	4.0	4.6	4.0	2.6	0.8	2.5	0.7
1986	4.6	4.0	4.5	4.1	2.1	1.1	2.1	1.0
1987	5.0	4.3	4.9	4.3	2.6	0.9	2.3	0.9
1988	5.1	4.2	4.9	4.2	2.4	0.9	2.4	0.8
1988b	5.4	4.4	5.2	4.4	2.6	1.0	2.6	1.0
1989	7.1	6.2	6.6	6.2	3.5	1.4	3.1	1.4
1990	5.4	4.6	5.4	4.6	5.3	4.6	5.4	4.6
1991	5.5	4.6	5.1	4.7	5.5	4.6	5.1	4.7
1992	5.8	4.9	5.4	4.9	5.7	4.8	5.3	4.8
1993	6.3	5.4	6.3	5.5	6.2	5.3	6.1	5.3
1994 ^b	6.6	5.4	6.3	5.4	6.6	5.2	6.2	5.3
1995 ^b	6.9	5.5	6.3	5.5	6.6	5.3	6.0	5.3
1996 ^b	6.7	5.8	6.4	5.9	6.4	5.5	6.1	5.6
1997 ^b	7.0	5.7	6.3	5.7	6.6	5.4	6.0	5.5
1998 ^b	6.9	5.7	6.5	5.7	6.7	5.5	6.3	5.6

Source: Author's calculations based on data from the California Department of Finance, the U.S. Census Bureau, and the March CPS 1970-1998.

^a Final weights are not provided in the March CPS before 1976. Using either the AWGT or PWGT in 1970-1975 leads to sums of the order of twenty million. In 1976-1988, the March weight is used in place of the final weight at the national level for persons under 14 years of age.

^b Beginning in 1994, the official national weights included estimates of the undercount population.

Notes: The table shows the sum of the absolute difference between the CPS sample and the independent population estimates across demographic groups defined by sex, age, and race/Hispanic origin.

Regional Distribution

Table 3.3 shows the population size by region using the California weights. For the Department of Finance population estimates, the California weights do not generally improve the CPS regional population counts. The one notable improvement is that the San Diego population was 4 percent too small using the national final weight and is correct using the California final weight (compare Table 2.2 with Table 3.3). For the Census Bureau population estimates, the California weights improve the CPS regional population counts by 1 to 4 percentage points in nearly every cell. The improvement is small, however, considering that the national weights provided a fairly close regional match (Table 2.2).

**Table 3.3. Independent and CPS Population Counts in 1998
by Region within California, California Weights**

	<u>Dept. of Finance</u>			<u>Census Bureau</u>		
	Estimate (1000s) ^a	CA Final weight (% diff)	CA March weight (% diff)	Estimate (1000s) ^a	CA Final weight (% diff)	CA March weight (% diff)
Los Angeles area	16,003	1	1	15,535	2	2
San Francisco area	6,628	0	1	6,453	0	0
San Diego area	2,676	0	0	2,631	-1	-1
Sacramento area	1,663	3	3	1,654	0	0

Source: Author's calculations based on data from the California Department of Finance, the U.S. Census Bureau, the 1990 decennial Census, and the 1998 March CPS.

^a Adjusted to reflect civilian noninstitutional population by calculating the percent in the Armed Forces and/or institutionalized from the 1990 decennial Census for each geographic region and applying that percent to the 1998 population estimate.

Notes: The Los Angeles area includes the counties of Los Angeles, Orange, Riverside, San Bernardino, and Ventura. The San Francisco region includes the counties of Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano, and Sonoma. The San Diego area is San Diego County. The Sacramento area includes Sacramento, Placer, El Dorado, and Yolo counties.

Social Indicators

Compared to the social indicators measured in the 1990 Census, the California weights make no improvement relative to the national weights. The poverty rate, distribution of household income, and distribution of education in the California subsample of the 1990 CPS remains essentially unchanged if one were to switch from national weights (Table 2.3) to California weights (Table 3.4). If anything, there is a very slight worsening of the match with the 1990 Census for a few cells (e.g., the poverty rate goes from 0.3 to 0.4 percentage points higher than the Census). For the most part, the use of California weights neither distorts nor improves the measurement of social indicators in the CPS.

Table 3.4. Social Indicators in the 1990 Census and CPS, California Weights

	Census (%)	Dept. of Finance		Census Bureau	
		CA Final weight (%)	CA March weight (%)	CA Final weight (%)	CA March weight (%)
POVERTY RATE ^a	12.6	12.9	12.9	13.0	13.0
HOUSHOLD INCOME ^a					
Less than \$20,000	20	22	22	22	22
\$20,000 - \$39,999	26	27	27	27	27
\$40,000 - \$59,999	22	23	23	23	23
\$60,000 - \$99,999	21	20	20	20	20
\$100,000 and over	11	9	9	9	9
EDUCATION ^b					
Less than H.S. diploma	20	18	18	18	18
H.S. diploma	20	28	28	28	28
Some college	34	27	26	27	26
Bachelor's	18	17	17	17	17
Advanced degree	9	11	11	11	11
Twelve years (adjusted)	24	28	28	28	28

Source: Author's calculations based on data from the 1990 decennial Census and the 1990 March CPS.

^a The 1990 surveys ask about income in 1989 so the poverty and income measures reflect 1989 levels.

^b Educational attainment reported for persons aged 25 to 50. The education distribution in the Census differs from that of the CPS in part because of a difference in the questions asked. In 1990 the Census asked about completion of degrees while the CPS asked about years of schooling. In the final row, "Twelve years" adjusts Census data to include persons finishing twelve years as well as those with a high school diploma.

Notes: Percentages may not add to 100 due to rounding.

I also investigated the social indicators in the 1998 March CPS. For poverty, the distribution of household income, and the distribution of education, the national and California weights provided essentially the same results (see Table 3.5). Similar analysis for 1976 and 1986 (not shown) showed that the distributions of these social indicators were not substantially changed when the California weights were used in place of the national March weights. For those years, the national final weights suggested slightly lower poverty and slightly higher household income.

The California weights are potentially more important when evaluating trends over past decades. The methodology for the CPS weights has changed several times since 1970. Furthermore, following each decennial Census, the CPS weighting algorithm was revised to

Table 3.5. Social Indicators in the 1998 CPS, National and California Weights

	<u>National Weights</u>		<u>Dept. of Finance</u>		<u>Census Bureau</u>	
	Final weight (%)	March weight (%)	CA Final weight (%)	CA March weight (%)	CA Final weight (%)	CA March weight (%)
POVERTY RATE ^a	16.9	16.6	16.5	16.6	16.6	16.7
HOUSHOLD INCOME ^a						
Less than \$20,000	19	19	19	19	19	19
\$20,000 - \$39,999	23	23	23	23	23	23
\$40,000 - \$59,999	17	18	18	18	18	18
\$60,000 - \$99,999	23	22	23	22	22	22
\$100,000 and over	18	18	18	18	18	18
EDUCATION ^b						
Less than H.S. diploma	19	19	18	18	19	19
H.S. diploma	23	24	24	24	24	24
Some college	30	30	30	30	30	30
Bachelor's	20	20	20	20	20	20
Advanced degree	8	7	8	8	7	7

Source: Author's calculations based on data from the 1998 March CPS with California weights.

^a The 1998 surveys ask about income in 1997 so the poverty and income measures reflect 1997 levels.

^b Educational attainment reported for persons aged 25 to 50.

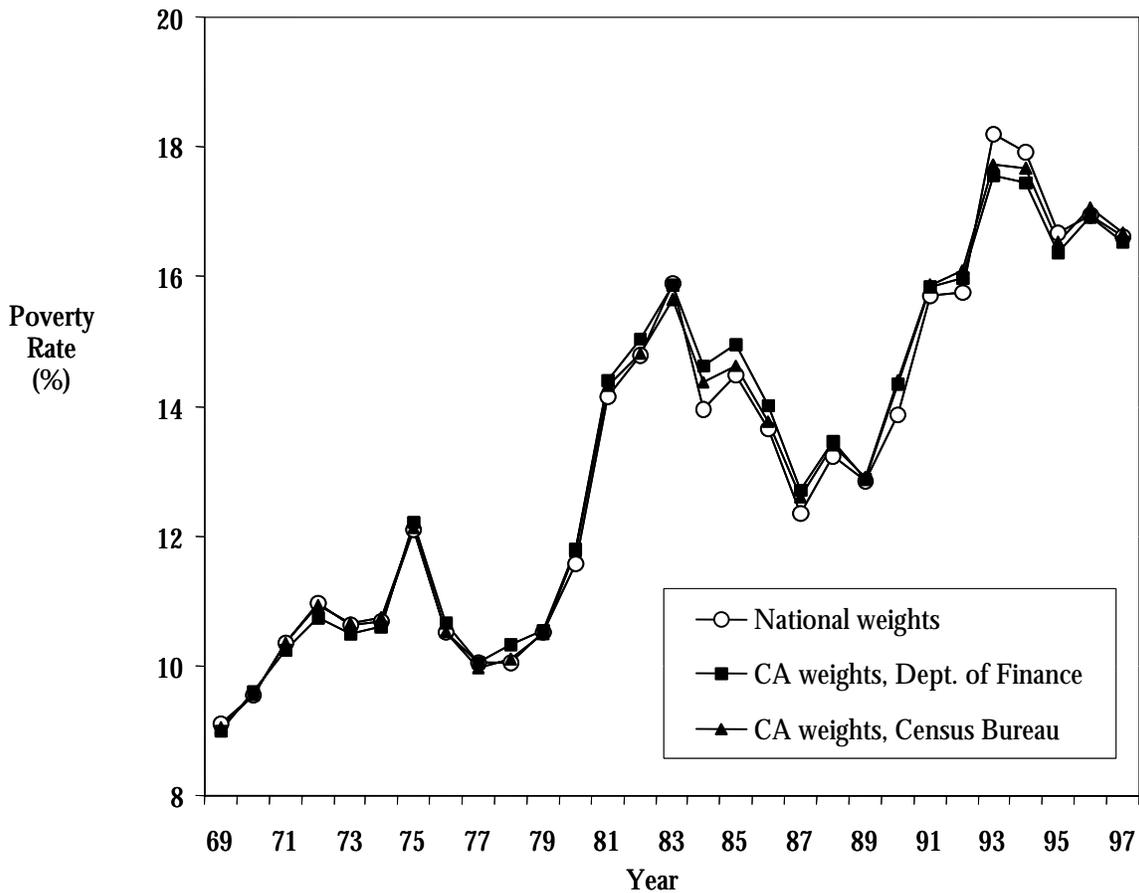
Notes: Percentages may not add to 100 due to rounding.

reflect new population estimates. The revisions in the weighting algorithm could potentially lead to jumps in the measurement of social indicators. In contrast, the California weights were developed by applying current weighting methods to all previous years. The California weights are based on independent population estimates that have been revised to incorporate Census information retrospectively, resulting in a more smooth transition of population counts.

Figure 3.1 shows the poverty rate in California as measured by the CPS using the national weights and the California weights. For most years, the choice of weights does not make a substantial difference in the estimated poverty rate. The difference is more notable for poverty in 1984 and 1993. Poverty for these years is measured by the CPS in March of 1985 and 1994, respectively. In both these years, revisions to the sample design and weighting algorithm were introduced. According to CPS technical documentation, the revisions affected the measurement of poverty. That is, due to the revisions, the poverty change from 1983 to 1984 and from 1992 to 1993 may not be accurate. Using the California weights diminishes the estimated poverty change in these years. This result suggests that the California weights may be more accurate than the national weights for measuring trends in

poverty. I also found less change in median family income in California in 1984 and 1993 using the California weights (not shown).

This chapter demonstrates the use of California weights that more closely match the distribution of sex, age, and race/Hispanic origin in the CPS to that of independent population estimates. The California weights slightly improve the regional population estimates in the CPS in 1998. However, the distributions of social indicators in 1990 and 1998 remain essentially unchanged. The California weights appear to moderately reduce the impact of survey redesigns in 1985 and 1994 on the estimates of poverty and income trends.



Source: Author's calculations based on data from the March CPS 1970-1998 for poverty in years 1969-1997.

Figure 3.1 -- Trends in the California Poverty Rate, National and California March Weights

4. Construction of the California Weights

In constructing the California weights, I followed the population-based weighting method used for the national weights in the CPS.¹⁶ In this chapter, I begin with a brief description of the multistage national weighting process. I then describe the seven basic steps for population-based weighting. Finally, I provide the details for each of the seven steps.

The process for creating national weights has several stages. The first stage adjusts for sampling probabilities. The second stage adjusts for non-interview (i.e., when an eligible, sampled household does not respond). The third stage adjusts for the sampling of geographic areas (often groups of counties, known as “Primary Sampling Units” or PSUs). In this step, the racial make-up (black versus non-black) of the sampled PSUs is adjusted to match the racial make-up in the state of residence. The final stage adjusts the sex-age-race-Hispanic origin groups to match independent estimates of the national population.¹⁷

To create the California weights, I repeated the final stage of the CPS process – adjusting the sex-age-race-Hispanic origin groups to match independent estimates of the state population for the subsample in California. I did not attempt to duplicate the first three stages because they require information on sampling, interviews, and PSUs not available in the public use survey. However, I retained the CPS estimates of the first three stages by using the national March weight as the starting point.¹⁸ Thus, individuals within the same sex-age-race-Hispanic origin group will have different California weights due to adjustments made in the first three stages of CPS national weighting.

I created two sets of California weights based on two different estimates of the California population: one from the California Department of Finance and the other from the U.S. Census Bureau. The process of creating the weights is summarized in the following seven steps:

1. Obtain independent estimates of the California population by sex, age, and race/Hispanic origin from the California Department of Finance and the United States Census Bureau.
2. Adjust the independent population estimates to remove Armed Forces and institutional populations not sampled in the CPS. Rates of Armed Forces and institutionalization were calculated at the national level from the decennial Census by sex, age, and race/Hispanic origin.

¹⁶ See *The Current Population Survey: Design and Methodology* (1978) as well as the *Technical Documentation* for each survey year.

¹⁷ Because of the rotating nature of the CPS sample (i.e., sampling by rotation groups), the Census Bureau constructs weights separately for each rotation group. In constructing the California weights, I ignored rotation groups.

¹⁸ The national final weight would make a better starting point because the March weight uses a final adjustment to be consistent within families (as described later in this chapter). I elected to use the March weight because the final weight is not provided before 1976, is not estimated for persons under age 14 until 1989, and is not provided for people in the Hispanic supplemental sample.

3. Collapse the independent estimates into aggregated age cells due to the small sample size of the CPS in California. Limit the race and Hispanic origin classifications to those identified in the CPS.
4. Calculate the sum of the March weights in each sex, age, and race/Hispanic origin cell and merge with the independent population estimates.
5. Iterate through the data ten times repeating two stages. The first stage creates a multiplier ratio so that within each race the distribution of sex and aggregated age matches the independent population estimates. The second stage adjusts the multiplier ratio so that within collapsed race groups (non-Hispanic and Hispanic) the distribution of sex and disaggregated age matches the independent population estimates. The purpose of the iteration is to create a single multiplier ratio that closely matches the criteria for both stages.¹⁹
6. The final weight is the product of the multiplier and the national March weight.
7. March weights require several further steps to ensure that families have consistent weights.

I describe below each of the seven steps in greater detail.

1. Obtain independent population estimates.

From the California Department of Finance I obtained estimates of the state population by sex and single year of age (up to 100) within the following race/Hispanic origin categories: white non-Hispanic, Hispanic, Asian non-Hispanic, black non-Hispanic, and Native American non-Hispanic. Estimates are available from 1970-1996. For 1997 and later, population projections are available.²⁰

From the United States Census Bureau²¹ I obtained estimates of the state population by grouped age for 1970-1980.²² For 1981-1989, the state population estimates are provided by sex and five-year age grouping (e.g., 0-4, 5-9, ... 80-84, 85+) within four race groups (white, black, Asian, Native American) where each of the race groups is subdivided into Hispanic and non-Hispanic. For 1990-1998, the estimates are provided by sex and single year of age within all eight race/Hispanic origin groups.

The population estimates from the Department of Finance and from the Census Bureau do not attempt to adjust for Census "undercount." However, undercount adjustments were incorporated in the CPS weighting algorithm beginning in 1994. For this reason, the California weights based on Census Bureau population estimates have lower population totals than the CPS weights in 1998 (see Table 3.1). Although undercount adjustments can lead to more accurate population representation, creating historic undercounts back to 1970

¹⁹ The national weights are created using a three-stage iteration. For further information see the detailed description of step 5 and footnote 24.

²⁰ See the Department of Finance website at <http://www.dof.ca.gov/html/Demograp/race.htm>. The counts reflect the population on July 1 of each year.

²¹ See the U.S. Census Bureau website at <http://www.census.gov/population/www/estimates/statepop.html>. The reference date for these estimates is July 1 of each year.

²² The age groups are 0-2, 3-4, 5-13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25-29, 30-34, 35-44, 45-54, 55-59, 60-61, 62-64 male, 62-64 female, and 65 and older.

was beyond the scope of this paper. When the 2000 decennial Census data are available, I intend to evaluate the importance of undercount adjustments for the California weights over the 1990s.

2. Adjust estimates for Armed Forces and institutional population.

The CPS samples the civilian noninstitutional population. From the decennial Censuses of 1970, 1980, and 1990, I constructed the Armed Forces and institutionalized rates at the national level²³ for each of the population cells available in the independent population estimates described above.²⁴ I then applied the Census rates to the population estimates to create counts of the civilian noninstitutional population in California by sex, age, and race/Hispanic origin. I used a linear interpolation to approximate weights for years between Censuses (and linear extrapolation after 1990²⁵).

The Armed Forces and institutional adjustments do not attempt to adjust for California-specific rates. Statewide independent estimates of these populations by sex, age, and race/Hispanic origin are not readily available. An alternative is to develop state-specific rates from the decennial Censuses. Due to sample size, the state-specific rates would require collapsing several age groups.²⁶ When the 2000 decennial Census data are available, I intend to explore the robustness of the California weights to alternative adjustments for the 1990s.

3. Collapse age and race/Hispanic origin groups

For the California weights, I used aggregated age groups due to the small sample size of the CPS in the state. I also needed to limit the race and Hispanic origin groups to those identified in the CPS.

For the weights based on the Department of Finance population estimates, I used the following age categories in all years: 0-2, 3-5, 6-8, 9-11, 12-14, 15-17, 18-20, 21-24, 25-29, 30-34, 35-39, ... 60-64, 65-74, and 75 and older. In 1970, the CPS did not identify Hispanics, so the California weights were created without controls for race or Hispanic origin.²⁷ Before 1988, the CPS did not identify Asians. For 1971-1988, I used four race/Hispanic origin groups: white non-Hispanic, Hispanic, black non-Hispanic, and "other" non-Hispanic. After 1988, I used all five race/Hispanic origin categories available from the Department of Finance.

For weights based on the Census Bureau population estimates, I used all 23 age cells available from 1970-1980. For 1981-1988, I used the five-year age categories available from

²³I excluded Alaska and Hawaii in making these estimates.

²⁴ For the Census Bureau population estimates beginning in 1981 I collapsed the race/ethnicity categories into white non-Hispanic, Hispanic, Asian non-Hispanic, black non-Hispanic, and Native American non-Hispanic. For the Department of Finance population estimates I collapsed the older age categories into 90 years and above.

²⁵ Exponential extrapolation led to rates that were too high for some groups. When linear extrapolation led to rates less than zero I used zero.

²⁶ The 1970 Public Use Sample is too small for state-based rates.

²⁷ In theory it would be possible to create race groups without distinguishing Hispanic origin. However, in the CPS survey, many white Hispanics identify as race "other" while others identify as race "white." Therefore, it is not possible to properly distinguish races without information on Hispanic origin.

the Census Bureau and the following race/Hispanic origin categories: white non-Hispanic, Hispanic, black non-Hispanic, and “other” non-Hispanic. For 1989, I used the five-year age categories with the following race/Hispanic origin categories: white non-Hispanic, Hispanic, Asian non-Hispanic, black non-Hispanic, and “other” non-Hispanic. For 1990-1998, I used the same age and race/Hispanic origin categories as with the Department of Finance weights.

4. Sum the March weights in the CPS for each group.

Steps 1 to 3 develop the “true” population estimates for each of the main sex, age, and race/Hispanic origin groups. The next step is to determine the number of persons in each group in California based on the CPS national March weights. For each year 1970-1998, I summed the March weights of the CPS for each of the cells described in Step 3 (above). I merged these sums to the population count data developed in the prior steps.

5. Iterate to calculate multiplier ratios.

The CPS sample for California has roughly 12,500 to 16,000 observations per year. Even for the collapsed groups described in Step 3, some of the groups have no sample observations. In this step, I created multiplier ratios to go from national weights to California weights accounting for the small sample size.

One option is to calculate the ratio of the “true” population to the CPS sample population for each group represented in the sample. This ratio could then be multiplied by each individual’s national March weight to create a final weight. However, the final distributions of age, of sex, and of race/Hispanic origin would not be correct because some groups are not represented in the sample. For example, in the 1998 sample there were no Native American girls aged 15 to 17 in the CPS. The “true” population estimate for this group was about 4,000. Therefore, using the simple ratio method, the total number of girls, the total number of Native Americans, and the total number of youth aged 15 to 17 would be 4,000 people too low.

One solution is to aggregate cells until every cell has at least some minimum number of sample observations. However, this method could lead to unacceptably high weights for small groups. For example, there are seven observations for black males aged 21 to 24 in 1998. For these seven observations to represent the “true” number of about 64,000, the average weight for the group would have to be over 9,000. By comparison, the average March weight is about 2,000. I chose not to use this method because it requires making a subjective decision regarding the minimum number of sample observations required for each group and/or the maximum average group weight.

To create the California weights, I chose to use the iteration method used in creating the national CPS weights. The iteration method has two stages.²⁸ In the first stage, I calculated

²⁸ The national weighting procedure has three stages and six iterations. The first stage creates a ratio that adjusts state populations to match independent estimates of state populations for civilian noninstitutionalized persons aged 16 and over. The first stage is not necessary for the California weights. The second stage adjusts to national population estimates for 14 Hispanic age-sex groups and five non-Hispanic age-sex groups. The third stage adjusts to national population estimates for 66 white age-sex groups, 42 black age-sex groups, and ten “other” age-sex groups.

the ratio of the “true” population to the CPS survey population within each race by sex and aggregated age groups. For the years 1990-1998, I used the age groups as defined in Step 3 for white non-Hispanics. For Asian non-Hispanics, I used the following age groups: less than 18, 18-39, 40-64, and 65 and older. For black non-Hispanics, I used the following age groups: less than 18, 18-64, and 65 and older. For Native American non-Hispanics, I combined sexes for persons under 18 and then grouped those over 18 by sex.²⁹ There is no first stage ratio for Hispanics because sex and age adjustments for Hispanics are done separately in the second stage.

In the second stage, I adjust the ratio calculated in the first stage to match the sex and disaggregated age distribution. The second stage is applied separately for Hispanics and non-Hispanics. For 1998 for non-Hispanics, I used the age groups described in Step 3. For Hispanics, I needed to aggregate further due to small cell sizes. I used the following age groups: less than 9 years, 9-17, 18-24, 25-34, 35-44, 45-54, 55-64, and 65 and older.

I iterated through these two stages ten times.³⁰ The purpose of the iterations is to create a final multiplier ratio that, when multiplied by the national March weight, would lead to population counts that closely match the “true” populations by sex and aggregated age within each race/Hispanic origin group and by sex and disaggregated age for non-Hispanics.

6. Calculate the final weights using the multiplier ratios.

The final weights for California are simply the product of the multiplier ratio calculated in Step 5 and the individual’s national March weight.

7. Calculate the March weights for families.

March weights required additional steps. Using the final weights, the number of married men will not necessarily be equal to the number of married women. Since one of the main purposes of the March supplement is to investigate demographic trends such as marriage, the CPS also includes March weights that adjust final weights to be consistent within families. I followed the national CPS procedure to calculate March weights for California.

I began by repeating the first stage of the iteration process (described in Step 5), so that the race/Hispanic origin distribution matched the “true” population counts. In the following discussion, I use the term “adjusted weight” to refer to the product of the multiplier ratio from this final iteration times the individual’s national March weight. For

²⁹ For the weights based on the Department of Finance estimates, the groupings for each stage are as described in the text for 1989-1998. For 1971-1988, the Asian and Native American groups are combined into “other” and I used the same sex and aggregated age groupings that I used for Asians in the later years. For 1970, I did not use race/Hispanic origin categories. In that year, for each sex, I matched directly to the “true” population using the age groupings described in Step 3 (i.e., I did not need to iterate).

For the weights based on Census Bureau estimates, the groupings are as defined in the text for 1990-1998. For 1981-1989 the Census Bureau estimates have age aggregated into five-year groupings so I adjusted our age groups accordingly (e.g., less than 18 becomes less than 20, etc.). For 1981-1988 I combined Asians and Native Americans into “other.” For 1970-1980 I did not have sex or race/origin categories. In those years I matched directly to the “true” population using the 23 age groupings available in the Census Bureau population estimates described in Step 1 (i.e., I did not need to iterate).

³⁰ The Census Bureau uses six iterations for the national weights.

women aged 15 years and older,³¹ the California March weight is equal to the adjusted weight. For each husband, the California March weight is set equal to that of his wife.³²

To calculate March weights for unmarried male family heads, I first calculated the average ratio of wife's March weight to husband's adjusted weight for married males by sex, age, race/Hispanic origin, and labor force status.³³ I then multiplied that ratio by the adjusted weight of each individual unmarried male family head.

For other adult males (aged 15 years and older, not married, and not family heads), I first calculated the sum of adjusted weights for all adult males by age, sex, race/Hispanic origin and labor force status. I then calculated the sum of adjusted weights for males who were married and/or family heads for the same groups. The difference between these sums is the number of other adult males required so that the group population will match the "true" group population. The multiplier ratio for the March weight of other adult males is the ratio of the number required over the sum of adjusted weights of other adult males (by age, sex, race/Hispanic origin, and labor force status). The California March weight for an "other adult male" is the product of this ratio times his adjusted weight.³⁴

For children under age 15,³⁵ I first created a temporary weight equal to the California March weight of the female family head or spouse of head. If there is no female family head or spouse, the temporary weight is the California March weight of the male family head.³⁶ For each sex, age, and race/Hispanic origin group, I calculated the ratio of the sum of the adjusted weights over the sum of the temporary weights. The California March weight is the product of this ratio times the temporary weight.³⁷

At this point, I deviated from the process used for national weights. In some years, the process of calculating March weights for males who are married and/or family heads creates a situation in which the sum of March weights for other adult males would need to be negative in order for the sum of March weights to total the "true" population count. This tends to happen for small groups such as older blacks and Native Americans. I used a final step to correct this problem. If the California March weight for an individual was less than one-tenth of his California final weight, I separated his entire age, sex, race/Hispanic origin group and all of their family members. I then started the March weighting process from the beginning. I first assigned to each husband and wife pair a weighted average of their adjusted weights. On the first iteration of this correction, I used $(2/3 * \text{wife} + 1/3 * \text{husband})$. I then assigned new March weights to all unmarried females aged 15 and older, based on an adjustment ratio that matched the total for each age and race/Hispanic origin group to the "true" total for the group. I then assigned weights to unmarried male family heads, other males, and children

³¹ The Census Bureau uses age 14 and older.

³² The Census Bureau uses the wife's weight to determine the husband's weight for a number of reasons including that coverage ratios are better for females (i.e., the multiplier ratios in Step 5 are lower for females).

³³ The labor force status categories are unemployed, agricultural worker, non-agricultural worker, and not in the labor force.

³⁴ In calculating the national March weights, unmarried males are assigned their final weight.

³⁵ The Census Bureau uses children under age 14.

³⁶ In rare cases, there is neither a female nor male family head identified. In those cases, I use the individual's own adjusted weight.

³⁷ In rare cases where a child under 15 is married or a family head (or spouse), the California March weight is calculated using the algorithm for adults.

under age 15, following the usual process for March weights (as described above). If this iteration did not correct the problem, I then assigned each husband and wife pair a weight equal to $(.5 * \text{wife} + .5 * \text{husband})$. If the problem remained, I assigned $(1/3 * \text{wife} + 2/3 * \text{husband})$ to married couples. When necessary, a final iteration was used $(0.1 * \text{wife} + 0.9 * \text{husband})$. Following the final iteration, there were never any cases in which the California March weight was less than one-tenth of the California final weight.

I did not separately identify family and household weights. However, the family weight is simply the California March weight of the family head, and the household weight is the California March weight of the household head.

Final detail for members of the Armed Forces

The CPS does not sample the Armed Forces population. However, if a member of the Armed Forces lives with civilians in a sampled household, that person will be included in the March survey. The CPS documentation is not clear on how weights are calculated for members of the Armed Forces. In constructing the California final weights, I removed members of the Armed Forces in all of the steps through Step 5. In Step 6, I applied the multiplier ratio calculated for civilians by sex, age, and race/Hispanic origin.

For the California March weights, I used the adjusted weight for female members of the Armed Forces. For males who were married and/or family heads, I used the regular March procedure described in Step 7 above. For other males, I calculated the multiplier ratio based on civilians and then applied it to members of the Armed Forces.

5. Accessing, Using, and Updating the California Weights

This chapter describes the proper use of CPS and California weights, how to access the California weights from PPIC, and plans for updating the California weights.

When to Use Weights and Which Weights to Use

For national estimates, users should use the official national weight, even for observations in California. For Pacific Region estimates, users should also use the national weights for all observations. California weights should be used only for California state-level estimates.

It is beyond the scope of this project to determine whether the Department of Finance or the Census Bureau has more accurate population estimates for the 1990s. However, for the years before 1990, weights based on the Department of Finance population estimates are preferred because the data has much finer sex, age, and race/Hispanic origin detail.

Researchers should use the same guidelines as used for the national weights to determine whether to use the final or March weight. Regarding the national weights, the Census Bureau offers the following guidelines in the Technical Documentation (1998, pp. 2-6). “(The) final weight should be used when producing estimates from the basic CPS data The March supplement weight should be used for producing estimates from the March supplement data.”

The family weight is simply the March weight of the family head, and the household weight is the March weight of the household head. Using household and family relationship identifiers, users can construct California family and household weights from the California March weights. Family (household) weights should be used for statistics that describe families (households). For example, “In California, 16.4 percent of families have incomes below the poverty line.” In many cases, the researcher is interested in the distribution of a family-level variable across people. For example, “In California, 16.6 percent of people live in families with incomes below the poverty line.”³⁸ For family (household) statistics at the person level, the Census Bureau recommends using the sum of March person weights to represent the family (household).³⁹

Weights should be used for descriptive estimates including population counts, means, and distributions. However, when estimating an individual-level statistical model, in most cases weights should not be used (e.g., in an OLS regression).⁴⁰ In a statistical model, the outcome (i.e., dependent variable) should be determined by the modeled explanatory (i.e.,

³⁸ Poverty statistics in this paragraph are based on the California March weight using the Department of Finance population estimates.

³⁹ Person-based weighting is used for calculating the official poverty statistics. This information was confirmed in correspondence with Gregory Weyland at the U.S. Census Bureau. The alternative approach, using the family (household) weight multiplied by the number of family (household) members, is not officially used.

⁴⁰ The discussion of the use of weights in the text follows from DuMouchel and Duncan (1983).

independent) variables. In a properly specified model, the use of weights should not substantially change the estimated parameters. In cases where using weights leads to substantially different results, this suggests a specification error. Variables that determine the weights (e.g., sex, age, race/Hispanic origin, and region for the CPS) should be included as explanatory variables, perhaps with interactions or other non-linear specifications.

There are some statistical models in which weights should be used. First, weighting methods are used to correct for heteroskedasticity. However, the proper weights for this correction will be estimated based on the model and will not be the same as the final and March weights in the CPS. Second, in some data sets but not the CPS, the sampling is based on an endogenous variable (or “choice”). When the outcome of interest is related to the sampling, then weights must be used to correct for “choice-based sampling.”⁴¹

How to Access California Weights from PPIC

The California weights are available free of charge from PPIC by sending an email to the author at reed@ppic.org. This section describes the structure of the weights data and the process for merging with the CPS data.⁴² The weights data set is arranged as a space-delimited ASCII data set for each year, 1970-1999. At the time of this study, the Census Bureau had not yet made available the population estimates for 1999, so the 1999 file has a value of negative one (-1) for the weights based on Census Bureau data.

For years 1976 through 1999, each observation in the weights data has six variables in the following order. HHSEQ is a number that identifies the household.⁴³ PERID is a number that identifies the person within the household. WTFNLDF and WTMARDF are the California final weight and the California March weight based on the Department of Finance population estimates. WTFNLDB and WTMARDB are the California final and March weights based on the Census Bureau population estimates. To attach the weights to the March file of the CPS, sort the California subsample by HHSEQ and PERID and merge.

For the years 1970 to 1975, the March file of the CPS has no unique household identifier. For users who buy the March files from the Unicon Corporation, observations can be uniquely identified by Unicon variables. For other users, California weights for these years are available on a limited basis from the author. The weights files for 1970 to 1975 have the following variables: LINENO, AGE, SEX, WGT, WTFNLDF, WTMARDF, WTFNLDB, WTMARDB, and _HHID. To attach to the March CPS files, sort the California subsample by _HHID, LINENO, AGE, SEX, and WGT. Note that _HHID is a Unicon-created text variable. In SAS, it should be read in using a format command (e.g., length _HHID \$ 12) and an ampersand (&) should follow _HHID in the input statement.⁴⁴

⁴¹ For example, the Panel Survey of Income Dynamics (PSID) intentionally oversamples the poor. Statistical models of the determinants of poverty should use sample weights to adjust for the oversampling.

⁴² For researchers at PPIC, the California weights are already attached to the CPS data files. See the README text at Mammoth:/research/library/march-cps-v00 for more details.

⁴³ In the CPS files, the variable HHSEQ is called PPSEQNUM from 1976-1988 and PH-SEQ from 1988b-1999. The variable PERID is called PP-POS from 1976-1988 and PPPOS from 1988b-1999. The names HHSEQ and PERID are used by the Unicon Corporation.

⁴⁴ The length is ten in 1970 and 1971, six in 1972, and 12 in 1973, 1974, and 1975. For 1971, there is one household with no _HHID. Set this value to “999999999”.

The weights data sets do not separately identify family and household weights. However, the family weight is simply the California March weight of the family head, and the household weight is the California March weight of the household head.

Finally, if you use the California weights, I am interested in your feedback. I would like to know what uses researchers have for the California weights. In particular, at the time of this writing, I have not found an application for which use of the California weights leads to a substantially different conclusion than suggested by the national weights. I would like to be informed of any such findings.⁴⁵

Future Updates

This study has found that while California weights can improve the representation of the CPS for the state, the weights do not substantially change estimates of several social and economic trends. Furthermore, the California weights are less important in recent years when the sample is designed on a state basis, the CPS weighting algorithm controls for total state population, and the algorithm is relatively sophisticated in terms of population controls. Based on these findings, I will not be updating the California weights on an annual basis.

I intend to update this study following the release of the 2000 decennial Census microdata currently scheduled for the Spring of 2003. Based on the 2000 Census, the Department of Finance and the Census Bureau will revise their state population estimates for the 1990s. I will use these revised estimates to modify the California weights for the 1990s. I will also use the 2000 Census to compare social and economic indicators as I did for the 1990 Census in this study.⁴⁶

⁴⁵ Please contact the author at reed@ppic.org.

⁴⁶ As noted in Chapter 4, the 1990s update will consider alternative adjustments for Armed Forces and institutional populations as well as for Census "undercount."

Bibliography

DuMouchel, William, and Greg Duncan (1983), "Using Sample Survey Weights in Multiple Regression Analysis of Stratified Samples," *Journal of the American Statistical Association*, 78(383):535-543.

Reed, Deborah, Melissa Glenn Haber, and Laura Mameesh (1996), *The Distribution of Income in California*, Public Policy Institute of California, San Francisco.

State of California, Department of Finance (1998), *Race/Ethnic Population with Age and Sex Detail, 1970-2040*, Sacramento, California, December.

U.S. Census Bureau, Population Division, Population Distribution Branch (1999), *State Population Estimates*, Washington D.C., October.

U.S. Department of Commerce and Bureau of the Census (1978), *The Current Population Survey: Design and Methodology*, Washington D.C., January.