Longitudinal Analysis of Annual Income Dynamics in California

Paul Ong
R. Varisa Patraporn
Douglas Houston

Longitudinal Analysis of Annual Income Dynamics in California[1]

By
Paul M. Ong, R.Varisa Patraporn, and Douglas Houston

Department of Urban Planning
University of California, Los Angeles

March 30, 2007

Introduction

This paper examines the economic status of families with children in California using a longitudinal analysis of short run income dynamics using the March Current Population Survey (CPS). The analysis tracks the potential movement of families from one income category into another over two points in time to assess the extent of economic instability for families over the short run. This paper uses data from the 2000, 2001, 2002, 2003 and 2004 March CPS surveys. For purposes of this paper, we define families with children as a group of two or more people who reside together and who are related by birth, marriage or adoption and who have their own children under 18 years of age residing with them. According to the 2000 Census, there were approximately 4,536,000 families with children in California. Families with children make up 39% of total households in California and 61% of the total state population.

Income categories are determined by multiples of the Federal Poverty Line (FPL). We determine a family's relative poverty status by taking a family's income and dividing it by the poverty threshold, or FPL. As described below, we use family poverty status to assign them to income categories. The FPL provides a standard national measure tracking poverty over time, but may underestimate poverty, especially in urban areas with a high cost of living. The FPL was developed in the 1960's and is set at three times the "breadbasket," a term referring to a minimum acceptable level of food for a particular family size and composition, and is adjusted annually based on inflation. For example, the 1999 FPL for a family of four was an annual income of $17,029. Table 1provides the poverty threshold based on the number of persons in a family unit for 1999. Note that the census provides additional FPL amounts based on the number of related children under 18, but we use the weighted average poverty thresholds for each family size.

Table 1. Poverty Threshold, 1999

Family Persons	Poverty Threshold
1	$8,501
2	$10,869
3	$13,290
4	$17,029
5	$20,127
6	$22,727
7	$25,912
8	$28,967
9 or more	$34,417

Source: U.S. Census (http://www.census.gov/hhes/poverty/threshld.html)

For the analysis, the small sample size limits the number of FPL-based income categories to two. Families with a family income that is 50-249% of the FPL are classified as "Poor & Working Poor." Families with a family income that is 250-549% of the FPL are classified as "Middle Class."

This paper has three sections. Section 1 summarizes findings from previous studies on income dynamics. Section 2 includes a discussion about the methodology and challenges of using longitudinal CPS data to understand income dynamics. Section 3 reports the result from our analysis, which are based on comparing family income dynamics in CA to family income dynamics in the rest of the United States including Washington, DC.

Section 1. Income Dynamics in the United States and California

Previous studies using the Survey of Income and Program Participation (SIPP), Panel Survey of Income Dynamics (PSID) and the CPS have shown variable and increasing levels of instability in family income. The *Los Angeles Times'* analysis based on data from the PSID,[2] found that changes in annual income experienced by individual families have grown larger over the last three decades (Gosselin 2004). In addition, they found that after excluding the "most extreme fluctuations" that "annual income swings, up or down, for 68% of all U.S. families (Gosselin 2004). One reason for the increased income instability is an increase in earnings instability. A study using CPS data found that earnings instability today is higher than the late 1960s and early 1970s (Cameron and Tracy 1998). A study using longitudinal employment data from the California Employment Development Department (EDD) found substantial upward earnings mobility for individual workers, particularly for those started with the lowest wages and/or switched industries. Such longitudinal analysis provides important insight into income dynamics that contrasts with cross-sectional studies that have found that real wages are declining for California's workers (Dardia et al 2002).

Fluctuations in income have occurred regardless of income category. Using SIPP panel data, DeNavas-Walt et al. (2004), found that "households in the lowest income quintile in 1996, 38 percent were in a higher quintile in 1999; of those originally in the highest income quintile,

---

[2] The Times' figures are based on the Panel Study of Income Dynamics, which has followed the same 5,000 nationally representative families and their children for close to 40 years.

34% were in a lower quintile 3 years later." (p.2). Formby, Bishop and Kim (2002) found a high rate of upward mobility for low-income families especially in the mid- and late-1990s. The authors concluded that 30% of all low income families (i.e., those with incomes below twice the official poverty line) in 1997 were no longer in the low income population one year later" (Formby, Bishop, and Kim 2002).

Overall the literature on income dynamics has found sizable upward and downward movements between economic classes, with the relative numbers of affected workers, families and households increasing over the last decade. The causes for such income dynamics include both contextual and individual factors. Researchers have linked the long run causes of such changes in income categories to changes in the U.S. labor market and household composition, such as layoffs, change in marital status and the addition of children or older parents (Jones and Weinberg 2000; US Bureau of the Census 1995).

Section 2. CPS Matching

Several studies have used the CPS to create panel data usually from the CPS March file (Cameron and Tracy 1998; Duleep and Regets 1997; Jones and Weinberg 2000; Madrian and Lefgren 1999; Peracchi and Welch 1995). The CPS was designed primarily as a cross-sectional survey. However, the CPS data permits matching of individuals across surveys in adjacent years, allowing for for the construction of two-year panels. "Matching is possible because the CPS resamples households. Each month a new sample of households, called a 'rotation' is added to the overall sample. The samples in this rotation are interviewed for four consecutive months, 'retired' for eight months and then interviewed for four more months. At any given time there are eight different rotations in the overall CPS sample"(Demombynes 2002, 2).

Madrian and Lefgren (1999) describe some of the challenges in matching CPS respondents longitudinally:

> "Because there is some measurement error in both the variables
> used to identify individuals over time and in the characteristics of
> some individuals at any point in time, any procedure used to match
> CPS respondents has the possibility of generating incorrect
> matches and failing to generate potentially valid matches"

Specific challenges in matching CPS data include: non-response, mortality, migration, and recording errors. Non-responses include those housing units for which individuals are not available, could not be contacted, or refused to cooperate. Also, some individuals die between survey months and will not be matched from the first to the second survey. Some errors will take place because the same identification code will be given to a household and individual that is not actually the same household and/or individual that was interviewed the first time generating "false" positive matches.

In order to reduce the potential error of matching CPS data, we "cleaned" the data for families for which there was not a clear match based on standard family identification codes. This process helped us generate a more "valid" data set. The process of "cleaning" included

three major stages (Table 2). In the first stage, we identified those records in question for which the gender of the respondent matched and for which the difference in age was less than 1 year. In Stage 2, we used a more subjective criteria to visually review the remaining records in question. Where possible, we identified the best match based on gender, sex, and whether an individual was coded as living in the housing unit in the previous year. In the last stage, we sorted families by Family Sequence Number and kept the first record for the remaining cases in which the person-level data did not reveal a clear "match" for a housing unit.

Table 2. Details of Data Cleaning Process for California Families

	03-04		02-03		01-02	
	Original Count	Unique Records	Original Count	Unique Records	Original Count	Unique Records
Original Count	674	612	646	470	860	477
Stage 1 Cleaning	62	27	176	69	383	149
Stage 2 Cleaning	8	2	38	6	85	22
Not Same Residence				-3		-3
Stage 3 Cleaning	2	1	16	8	28	14
Total "Valid"		**642**		**553**		**662**

To assemble panel data of families with children from 3 years of CPS data (each based on two adjacent years of respondents), we assembled three data sets that with linked responses for three years: 2001-2002, 2002-2003, and 2003-2004. Each data set was limited to families with children (families with own or related children under 18). Also, we limited each annual data set to the "primary family" in a housing unit. We also only kept linked family record pairs for families for which the household was not a different "replacement" family in the second year. Finally, we appended the 3 separate annual files for the analysis. For a more detailed description of the data assembly please refer to Appendix A.

Despite efforts to develop the most "valid" matches for a longitudinal CPS panel data set, the analysis could still be biased by reporting errors that are not directly related to linking records across years. Simple cross-sectional analysis using a single year's of CPS assumes that the mean of such reporting errors is zero. That is, the over reporting of income offsets the under reporting of income in a given year. When matching two years of data, it is critical to consider how the error in one year is related to the error in the other year. Since each year of linked family records may have its own reporting error, the problem generated by misreporting may be compounded by linking family records longitudinally. Even with "good" matches in which the same family was interviewed in the two interviews, responses may vary in accuracy depending on the family member who responded and their familiarity with, or memory of, their family's income. Even if the same family member responds in two consecutive interviews, his or her response may vary in accuracy in each year depending on factors such as memory or perceived financial strain. There is no reason to believe that the reporting errors for income are identical in size and direction for both years. Therefore, the reporting errors for a given family may be compounded rather than offsetting when trying to estimate changes in income over the two-year period. For instance, over reporting in one year and under reporting in the subsequent year within a given family could produce artificial changes in income. Temporal errors make studying income dynamics with CPS data problematic and are likely to upwardly bias the estimates of the extent of movement between income categories.

4

Given this potential temporal reporting error, we do not report the absolute or percentage "movement" of families between FLP-based income categories. Rather, we assume that such errors occur at the same rate in data from California as they do in data from the rest of the nation and report. With this assumption, we can use the data to examine whether families in California are experiencing relatively more or less movement between income categories than families in rest of the nation. This approach allows us to use CPS longitudinal data to make useful comparisons about the relative income dynamics of California families with children.

Section 3. Analytical Results

Tables 3 presents the annual income dynamics of persons in families with children based on 3 years of "cleaned" panel CPS data.[3] It compares the change in income category of persons in California compared to persons in the rest of the nation. A score of less than 1.0 suggests that California had a lower rate of having persons in that category than the rest of the nation. A score of more than 1.0 suggests that California had a higher rate of having persons in that category than the nation. For example, 0.98 means that California had a lower rate of individual's in poor and working poor families moving down an income category (to very poor) relative to the individual's in the rest of the nation in poor and working poor families.

Table 3. Income Category Change for California versus Rest of the U.S.

	Income Category Change From Year 1 to Year 2		
Income Category Year 1	Moved "Down"	Same	Moved "Up"
Poor & Working Poor (FPL 50-249)	0.98	1.05	0.87
Middle Class (FPL 250-549)	1.21	0.88	1.22

Source: U.S. Census Bureau, Current Population Survey, March 2000-2004.

A few general trends hold when comparing California income dynamics to those of the nation:

- Persons in poor & working poor families in California had a higher rate of staying in this category than the same persons in the rest of the nation (1.05). Those in California also had a lower rate of moving into a higher income category than those in the rest of the nation (0.87).

- Persons in middle class families in California had a lower rate of staying in this category than the same persons in the rest of the nation (0.88). Those in California also had a higher rate of moving into a lower income category or into a higher income category than those in the rest of the nation (1.21 and 1.22).

Overall, persons in middle class families in California appear to have had a higher rate of change up or down compared to those in the rest of the nation, indicating a higher likelihood of increased income inequality. Poor and working poor families in California had a lower rate of change compared to the rest of the nation, indicating fewer opportunities for upward mobility in this state compared to the rest of the nation.

---

[3] The results presented in Table 3 are consistent with results based on different data assumptions. Appendix B presents results using "uncleaned" data and more conservative assumptions. Appendix A provides detailed information on the process of "cleaning" the data.

Appendix A. Current Population Survey (CPS) Data, Sample and Assembly

CPS Data

The CPS is a survey of housing units, not households, and starting in 1968 the Census Bureau has maintained unique residence identifiers, which are common across interviews.

The Structure of CPS Panel Data

We assembled a panel data set of families with children by drawing from 3 years of Current Population Survey (CPS) data. CPS was designed to be used primarily for cross-sectional data analysis in a given month, or for comparison of cross-section analysis at different points in time. Longitudinal analysis of housing units using data from consecutive CPS years is possible given that about 50% of the CPS sample is common from one year to the next for the same month. The use of CPS data as a longitudinal dataset, however, is limited since the short duration of the linked data limits its use for analysis of long-term trends. Also, since the CPS was not primarily designed as a panel data set, matching individuals or families across annual CPS data sets is not straightforward (Madrian and Lefgren 1999). This is in part to the fact that the CPS is a survey of housing units, not individuals or families, and therefore second interviews of a rotation group often contain responses from individuals in families who recently moved into the housing unit. Also, multiple families could reside in a single housing unit. Finally, coding errors often frustrate efforts to validate if the same individual or family responded in the second interview.

The following text by Madrian and Lefgren (1999) below describes the organization of CPS panel data:

> "The CPS is a monthly survey of a probability sample of housing units. It does not, however, survey a completely new set of housing units each month. Rather, the sample is divided into eight representative subsamples called rotation groups, with housing units in each rotation group being interviewed for four consecutive months, followed by an 8-month break, and then by another four months of interviews. Thus, CPS sample housing units are each eligible for 8 different monthly interviews, and rotation groups are referred to in CPS parlance by their "month in sample" or MIS. In any given monthly sample, approximately one-eighth of sample units will be interviewed for the first time (MIS=1), one-eighth for the second time (MIS=2), and so on. One-eighth of the sample will be leaving the sample permanently (MIS=8), and one-eighth will be leaving for the next eight months before being reinterviewed (MIS=4). These latter two rotation groups, MIS=8 and MIS=4, are referred to as the "outgoing rotation groups."

> "…Rotation group A is first interviewed in January of year *t* (MIS=1). It is subsequently interviewed in February, March and April of the same year (MIS=2, 3 and 4 respectively). Following this fourth interview in April of year *t*, rotation group A then leaves the sample for 8 months and is next interviewed in January of year *t+1* (MIS=5). It continues to be interviewed in February, March and April of year *t+1* (MIS=6, 7 and 8), and then leaves the sample permanently after the 8th and final interview in April of year *t+1*. The January *t+1* sample of the CPS is comprised in part of rotation group A

(MIS=5), along with rotation groups M, L, K and J (MIS=1, 2, 3 and 4 respectively), and rotation groups z, y and x (MIS=6, 7 and 8)."

The Matching Process

To assemble panel data of families with children by drawing from 3 years of CPS data, we first assembled three data sets that with linked responses for three years: 2001-2002, 2002-2003, and 2003-2004. To merge for each annual rotation group, we extracted a data set from *year 1* for MIS in (1 2 3 4) and a data set from *year 2* for MIS in (5 6 7 8) since these months represent the months in which the same rotation group was surveyed. Each data set was limited to families with children (families with own or related children under 18 [FOWNU18>0 or FRELU18>0]). Also, we limited each annual data set to the "primary family" in a housing unit [FAM_TYPE=1]. We also only kept linked family record pairs for families for which HHNUM=1, meaning the household was not a different "replacement" family in the second year.

We merged data for *year 1* and *year 2* based on the HHIDNUM, keeping only those families (based on HHIDNUM) which appeared in both years for the specified months and which were identified as a primary family. We then kept only linked interview pairs which were interviewed in the appropriate months. That is, interviews from *year 1* for MIS 1 were matched with interviews from *year 2* for MIS 5. Interviews from *year 1* for MIS 2 were matched with interviews from *year 2* for MIS 6, Etc.

This merge resulted in duplicate records per the housing unit identifier. That is, for a given housing unit, the merged data contained multiple records for each housing unit for a given month-pair and for a primary family. This could be due to responses from different individuals, or due to coding errors when identifying family characteristics such as which responses were assigned to family.

The existence of duplicate records per family could introduce error into tracking family income trends since some records likely represent different individuals or families with potentially very different income patterns. Therefore, we used the following screening procedures to process duplicate linked pairs in order to select the linked pair records that most likely represent the same individual, and therefore the same family in the second time period. Except in a few circumstances in which a family was coded as not living in the same housing unit in the previous year, our approach was to keep one linked record for all "valid" family records (using the criteria identified above).

To identify the "best" linked pair for HHIDNUMs with duplicates, we extracted person-level data and merged it to family data by the Family Sequence Number which is unique per year, but is not the same across years. First, we kept all families records (based on HHIDNUM) *without* duplicate linked records. Second, we screened the linked family pairs with duplicate records per HHIDNUM using the following stages (See Table A1 for details on the number of records processed at each stage per year):

- Stage 1 Cleaning: We identified those records for which the gender of the respondent matched and for which the difference in age was less than 1 year. In cases in which a given HHIDNUM had a total of 2 duplicates and for which only one individual matched on both criteria, we kept this "best" match assuming that the same individual matched on both years.

- Stage 2 Cleaning: Using more subjective criteria, we visually reviewed remaining duplicates and where possible, identified the best match based on gender, sex, and whether an individual was coded as living in the housing unit in the previous year. In a small number of cases, we eliminated families who did not have a clear match in Stage 1 and in which all individuals in the family reported they did not live in the same housing unit in the previous year. In this stage, we did not eliminate duplicates for which there was not a visible elimination criteria.

- Stage 3 Cleaning: For the remaining cases in which the person-level data did not reveal a clear "match" for a housing unit, we sorted families by Family Sequence Number and kept the first record.

For the analysis, the small sample size limits the number of FPL-based income categories to two. Families with a family income that is 50-249% of the FPL are classified as "Poor & Working Poor." Families with a family income that is 250-549% of the FPL are classified as "Middle Class."

Table A1. California: Details of Data Cleaning Process

	03-04		02-03		01-02	
	Original Count	Unique Records	Original Count	Unique Records	Original Count	Unique Records
Original Count	674	612	646	470	860	477
Stage 1 Cleaning	62	27	176	69	383	149
Stage 2 Cleaning	8	2	38	6	85	22
Not Same Residence				-3		-3
Stage 3 Cleaning	2	1	16	8	28	14
Total "Valid"		**642**		**553**		**662**

Source: U.S. Census Bureau, Current Population Survey, March 2000-2004.

Table A2. Nation: Details of Data Cleaning Process

	03-04		02-03		01-02	
	Original Count	Unique Records	Original Count	Unique Records	Original Count	Unique Records
Original Count	7379	6590	9172	5575	10790	5431
Stage 1 Cleaning	789	340	3597	1428	5359	2044
Stage 2 Cleaning	109	16	741	121	1271	146
Not Same Residence		-1		-10		-50
Stage 3 Cleaning	64	32	366	181	777	383
Total "Valid"		**6,978**		**7,305**		**8,004**

Source: U.S. Census Bureau, Current Population Survey, March 2000-2004.

Appendix B. Detailed Tables: Income Category from Year 1 to Year 2 for Poor, Working Poor and Middle Class Families

This appendix replicates the results presented in Table 3 of the main text based on different data assumptions, and demonstrates that the results "hold" across various data assumptions. First, Tables B2 and B3 present results using "uncleaned" data. Second, Tables B1 and B2 present a subset using "conservative" data assumptions. That is, to try to ensure that the same family member responded to the CPS survey in both time periods, we used a subset of families linked annual records in which the individual respondent was of the same sex and within 1 year of age in each time period.

Table B1. "Cleaned" Data with only Individual Respondents of the Same Sex and within 1 year of age.

	Income Category From Year 1 to Year 2		
Income Category Year 1	Moved "Down"	Same	Moved "Up"
Poor & Working Poor (FPL 50-249)	0.99	1.06	0.84
Middle Class (FPL 250-549)	1.25	0.88	1.25

Source: U.S. Census Bureau, Current Population Survey, March 2000-2004.

Table B2. Not "Cleaned" Data with only Individual Respondents of the Same Sex and within 1 year of age

	Income Category From Year 1 to Year 2		
Income Category Year 1	Moved "Down"	Same	Moved "Up"
Poor & Working Poor (FPL 50-249)	0.99	1.05	0.85
Middle Class (FPL 250-549)	1.24	0.88	1.24

Source: U.S. Census Bureau, Current Population Survey, March 2000-2004.

Table B3. "Not "Cleaned" Data

	Income Category From Year 1 to Year 2		
Income Category Year 1	Moved "Down"	Same	Moved "Up"
Poor & Working Poor (FPL 50-249)	1.00	1.06	0.87
Middle Class (FPL 250-549)	1.17	0.89	1.17

Source: U.S. Census Bureau, Current Population Survey, March 2000-2004.

References

Cameron, Stephen, and Joseph Tracy (1998) "Earnings Variability in the United States: An Examination Using Matched-CPS Data" Prepares for the 1998 Society of Labor Economics Conference.

Dardia, Michael, Elisa Barbour, Akhtar Khan, and Colleen Moore (2002) "Growth and Employment; Moving Up? Earnings Mobility in California." Public Policy Review, April 2002. Burlingame, CA: The SPHERE Institute and the Stanford Institute of Economic Policy Research.

DeNavas-Walt, Carmen, Bernadette D. Proctor, and Robert J. Mills (2004) "Income, Poverty, and Health Insurance Coverage in the United States: 2003" US Census Bureau: Current Population Reports.

Demombynes, Gabriel M (2002) "Three Ways of Looking at Immigrant Wage Growth: Analysis With the 1993-1998 Current Population Survey" Prepared for the Institute for Labor and Employment Conference, January 18-19, 2002

Duleep, Harriet Orcutt, and Mark C. Regets (1997) "Measuring Immigrant Wage Growth Using March CPS Files" *Demography* 34: 239-49

Formby, John P., John A. Bishop, and Hoseong Kim (2002) "The Economic Well-Being of Low-Income Working Families" University of Alabama: the Employment Policies Institute.

Gosselin, Peter (2004) "The New Deal" *Los Angeles Times,* 10 October 2004.

Jones, Arthur F, and Daniel H. Weinberg (2000) "The Changing Shape of the Nation's Income Distribution" US Census Bureau: Current Population Reports.

Madrian, Brigitte and Lars John Lefgren (1999) "A Note on Longitudinally on Matching Current Population Survey Respondents" Technical Working Paper 247, Cambridge, MA: National Bureau of Economic Research (http://www.nber.org/papers/T0247)

Peracchi, Franco, and Finis Welch (1995) "How Representative Are Matched Cross-Sections? Evidence From the Current Population Survey?" *Journal of Econometrics* 68:153-79

US Bureau of the Census (1995) "Income and Job Mobility in the Early 1990's" Statistical Brief, February, 1995.