The Earned Income Tax Credit: Its Impact on California's Low-Income and Welfare Populations

Final Report to the

State of California Franchise Tax Board

Submitted by

V. Joseph Hotz
Department of Economics
University of California, Los Angeles
Los Angeles, CA 90095
hotz@econ.ucla.edu

John Karl Scholz
Department of Economics and Institute for Research on Poverty
University of Wisconsin, Madison
Madison, WI 53706
jkscholz@facstaff.wisc.edu

September 22, 2005

Acknowledgements

This project was supported, in part, by contract number 95ASPE270A, granted by the Assistant Secretary for Planning and Evaluation of the U.S. Department of Health and Human Services (DHHS) and by the state of California Department of Social Services (CDSS) and NSF Grant No. 25-3385-01-0-79-461-7700 granted by the National Science Foundation. The investigators which to acknowledge the collaboration of two co-authors, Carolyn Hill and Charles Mullin on some of the reports included below, as well as the dissertation research completed by Eduardo Faynzylber under the supervision of Hotz that was completed as part of this project. We also wish to acknowledge the able research assistance of Matthew Cox, Carolyn Hill, Douglas Houston, Douglas McKee, Oscar Mitnik, Charles Mullin and Juan Pantano on various aspects of this project. We also wish to thank Jacob Klerman, Steven Haider, others associated with the RAND Statewide CalWORKs Evaluation, Lois Van Beers, Nikki Baumrind, Leslie Raderman, Paul Smilanick, Everett Haslett, Webb Hester, and others of the California Department of Social Services for their help in constructing some of the non-tax data used in this project, and Don Oellerich of the Office of the Assistant Secretary for Planning and Evaluation of the U.S. Department of Health and Human Services, Patricia Ruggles, formerly of ASPE and now at Georgetown University, and Phil Spilberg of the California Franchise Tax Board for their encouragement and support of this project.

Finally, we want to single out two people who deserve special recognition and thanks: Werner Schink, formerly of the California Department of Social Services, and George Ramsey, of the California Franchise Tax Board. Without Werner Schink's help and vision, this project would have never gotten off the ground. Without George Ramsey's guidance, patience and constant support, we never would have finished any of the work presented below. We owe each our eternal gratitude.

Contents of Final Report

Trends in EITC Take-Up and Receipt for California's Welfare Population, 1992-1999*

V. Joseph Hotz
UCLA

Charles Mullin
Vanderbilt University

John Karl Scholz
University of Wisconsin, Madison

August 30, 2003

*This research was funded, in part, by a grant from the Office of the Assistant Secretary for Policy and Evaluation at the U.S. Department of Health and Human Services. We wish to thank George Ramsey of the California Franchise Tax Board and his staff for their help with this project and Oscar Mitnik and Eduardo Fajnzylber for their able research assistance. All opinions expressed in this report and any errors are the sole responsibility of the authors. In particular, this paper does not necessarily represent the position of the U.S. Department of Health and Human Services, the State of California or its agencies.

SUMMARY

In this report, we examine the patterns and trends in EITC participation and amounts of credit received for adults on California's AFDC and adults on California's welfare (AFDC and TANF) caseloads for the period 1992-1999. Over the 1990s, EITC utilization rates and credits received by welfare recipient households in California increased, providing increasing levels of financial assistance to these households, in both absolute and relative (to labor market earnings) terms. Furthermore, our results indicate that an increasing fraction of households that appear to be eligible for the EITC actually claimed it over the 1990s. At the same time, our estimates suggest that sizeable numbers of California's working poor households who were eligible for the EITC did not claim it during the 1990s. We estimate that in 1999, only 73 percent of households in California that had been on welfare in the previous year but had wage and salary earnings in 1999 actually claimed the EITC on federal tax returns. Moreover, this participation rate was between 7 and 13 percentage points lower than for the U.S. as a whole. As a result, we estimate that in 1999 as many as 860,000 households in California who were eligible for the earned income credit did not receive it, foregoing as much as $1.4 billion in tax credits from the federal government.

1. Introduction

Since its inception in 1975, the Earned Income Tax Credit (EITC) has grown into the largest federally funded means-tested cash assistance program in the United States. In 2002, for example, some 19.6 million tax filers received $33.2 billion in earned income credits. In California, some 2.3 million filers received $3.9 billion in credits, accounting for roughly 12 percent of the Nation's filers and amount of credit received.

The continuing support for and expansion of the EITC has often been attributed to the perception that it is a public assistance program that rewards work. Only low-income households with adults who work are eligible for the credit. Moreover, recent studies have suggested that the expansion of the EITC over the last 15 years was a major contributor in the rise of employment and decline in welfare use among female-headed households.[1] The expansion of the EITC, especially during the 1990s, was accompanied by federal and state government efforts to reform their welfare programs, with these reforms also focused on encouraging work among disadvantaged families in order to encourage them towards greater financial self-sufficiency. In California, these reforms have been undertaken beginning with reforms of the State's past Aid to Families with Dependent Children (AFDC) program in the early 1990s and culminating in the California Work Opportunity and Responsibilities to Kids (CalWORKs) program, California's implementation of the Temporary Assistance to Needy Families (TANF) program that was enacted in 1998.

In this report, we examine the patterns and trends in EITC take-up and amounts of credit received for adults on California's AFDC and adults on California's welfare (AFDC and TANF) caseloads. More precisely, we analyze the incidence and amounts of earnings, and the incidence of tax filing and claiming of the EITC for adults in assistance units who were on AFDC or TANF in the fourth quarter of the year preceding the tax year. Because having wage and salary income is one of the necessary conditions for being eligible to claim the EITC, we present estimates both of the proportion of our samples that were employed (i.e., had wage earnings) during a particular tax year and of the average annual wage and salary earnings, since, as discussed below, the amount of the earned income credit one is eligible to receive depends on a household's earnings during a tax year. We use data on wages and salaries that are reported to California's Employment Development Department (EDD) for jobs covered by the State's Unemployment Insurance and Disability Insurance (UI/DI) system. We also present statistics on the incidence of federal tax filing in various years among the adults in our sample, since the EITC is provided to those who file tax returns.

In what follows, we provide a brief description of the structure of the EITC, describe the data we used in our analyses, present our findings and discuss some of their implications for welfare policy in the State.

2. Background on the EITC

The EITC is a refundable tax credit, meaning that it is paid out by the U.S. Treasury regardless of whether the taxpayer has any federal income tax liability. To receive the earned income credit, taxpayers file their regular tax return and fill out the six-line Schedule EIC that gathers information about qualifying children. To be eligible for the credit, a taxpayer must have both earned and adjusted gross income below a threshold that varies by year and by family size. Most

[1] See, for example, Meyer and Rosenbaum (2001) and Grogger (2003).

EITC payments go to taxpayers with at least one "qualifying child." A qualifying child needs to meet age, relationship, and residence tests. The age test requires the child to be younger than 19, younger than 24 if a full-time student, or any age if totally disabled. The relationship test requires the claimant to be the parent or the grandparent of the child or for the child to be a foster child.[2] Under the residence test the qualifying child must live with the taxpayer at least six months during the year.[3]

In 1999, taxpayers with two or more children could receive a credit of 40 percent of income up to $9,540, for a maximum credit of $3,816. Taxpayers (with two or more children) with earnings between $9,540 and $12,450 received the maximum credit. Their credit was reduced by 21.06 percent of earnings between $12,450 and $30,580.[4] Table 1 shows the changes in the "parameters" of the EITC, including the income eligibility thresholds, credit rates, and phase-out (or implicit tax) rates, for the 1992-1999, the period we analyze. As indicated in Table 1, a small credit available for childless taxpayers between the ages of 24 and 65 with very low incomes was added in 1994. As of 1999, the credit rate for these taxpayers is 7.65 percent, with a maximum credit of $347.

3. Data Used

Our analysis is based on annual cross-sectional samples of adults in AFDC/TANF assistance units that we refer to as annual "*Past Welfare Recipient"* samples, given that they consist of individuals who were on welfare in Quarter 4 in the year prior to the *tax year*[5] for which the outcome variables we analyze are recorded.[6] (The tax years analyzed are 1992 through 1999.)

The data on earnings are based on EDD earnings using the wage and salary earnings of adults in an assistance unit, where earnings are drawn from the EDD UI base wage file. Similarly, we aggregated all tax returns for individuals in these cases into a case level return. If anyone in a case files a return or claims the EITC, then the case filed a return or claimed the EITC. The size of the credit received by the case is the sum of the credits received by individuals in the case (joint returns, common in Unemployed Parent (UP) cases, were not double counted). Finally, all of the statistics reported in the tables in this report are weighted by the case-level weights.

Details of how the annual samples, sampling weights and variables analyzed were con-

---

[2] Until late 1999, a foster child was any child for whom the claimant cared for "as if the child is their own." The caring stipulation still holds, but now the child must also be placed in the home by an authorized placement agency. Prior to the 2001 tax legislation, EITC-eligible foster children also needed to live with the taxpayer for 12, rather than 6, months.

[3] In 1990 (tax year 1991) the residency and AGI tiebreaker (see below) tests replaced a support test, since in principle it is easier to verify where a child lives than it is to verify who supports a child. Under the support test the taxpayer had to pay for at least half the child's support, where items like transfer payments (i.e., AFDC and housing subsidies) and child support were not considered support provided by the taxpayer.

[4] In subsequent years, the maximum credit and thresholds for the Phase-in and Phase-out ranges increased.

[5] Tax Year refers to the year in which any earnings were realized. In most cases, tax returns on these earnings are filed in the following calendar year, usually by April 15th.

[6] It is important to note that the adults in these annual samples may have been on aid (on AFDC or TANF) in Tax Year *t*; our strategy for constructing these annual samples ensures is that they were on welfare in the last quarter of the year prior to the Tax Year.

structed are provided in the Appendix.

Estimates are presented below for six regions of the State as well as the entire State. These regions are those defined by the California Department of Social Services (CDSS)[7] and are displayed in Figure 1.[8] (Comparable statistics at the county level are available at http://www.econ.ucla.edu/hotz/Calif_EITC.) In Table 8, we provide a set of aggregate economic, poverty and public assistance program participation characteristics of these regions as of 1999.

In Table 2, we present the demographic characteristics of the annual Past Welfare Recipient samples by year and by region of the State.

4. Findings

4.1 Wage and Salary Earnings

We begin by examining the trends and cross-regional patterns in the fraction of households/assistance units in our past welfare recipient samples in which adult members had some wage and salary income during the tax years 1992-99. As already noted, having some wage and salary income during a tax year is a necessary, although not sufficient, condition for being eligible to claim the EITC.

The incidence of households/assistance units that had positive wage and salary earnings in a particular tax year increased substantially over the eight-year period, 1992-1999 (Table 3). For the state as a whole, the fraction of households with positive earnings for adults on welfare in the quarter immediately preceding a particular tax year went from 0.37 to 0.69, an increase of 86.5 percent. The trends in the incidence of households with wage and salary earnings differed substantially across of the regions of the State. In 1992, the incidence of adults in welfare assistance units with positive earnings was the lowest in Los Angeles County at 0.31, while highest the Northern and Mountain counties at 0.42. But, over the 1990s, the incidence of adults with at least some annual earnings grew more rapidly in Los Angeles County (118.5 percent) followed by the other southern California counties (94.1 percent). In contrast, the incidence of positive earnings households grew most slowly in the Northern and Mountain counties (58.8 percent) and the Central and Farm counties (67.6 percent) over the 1990s.

Over this same period, the average *amount* of wage and salary earnings of adults on welfare grew by almost 176 percent in real terms, from $2,273 to $6,271. In contrast, mean income for all families in the western states of the U.S. increased by only 15.3 percent over the same pe-

---

[7] See Research and Development Division & Data Analysis and Publications Branch, *The Regions of California: Recommended Grouping of the Counties for Regional Studies*, California Department of Social Services, August 2002, for a discussion of how this regional classification system was derived.

[8] The six regions, and the counties that comprise them, are: ***Bay Area*** (Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Santa Cruz, Sonoma counties); ***Southern California, excluding Los Angeles*** (Orange, Riverside, San Bernardino, San Diego, Santa Barbara, Ventura counties); ***Los Angeles county***; ***Central/Southern Farm*** (Fresno, Imperial, Kern, Kings, Madera, Merced, Monterey, San Benito, San Joaquin, San Lois Obispo, Stanislaus, Tulare counties); ***Northern and Mountain*** (Alpine, Butte, Calaveras, Del Notre, Glenn, Humboldt, Inyo, Lake, Lassen, Mariposa, Mendocino, Modoc, Mono, Nevada, Plumas, Shasta, Sierra, Siskiyou, Tehama, Trinity, Tuolumne counties); and ***Central Valley*** (Colusa, El Dorado, Placer, Sacramento, Sutter, Yolo, Yuba counties). See Research and Development Division & Data Analysis and Publications Branch, *The Regions of California: Recommended Grouping of the Counties for Regional Studies*, California Department of Social Services, August 2002.

riod.[9] As with employment, those on welfare in Los Angeles County experienced the most rapid earnings growth over this period, increasing by 213.5 percent in real terms from 1992 to 1999. Over this same period, the earnings of those previously on welfare also increased markedly in the Central Valley and non-farm southern counties, increasing by 186.4 and 182.5 percent, respectively. In contrast, the earnings of adults on welfare grew more slowly in the Northern and Mountain (120.4 percent) and Central and Southern farm (140.5 percent) counties.

4.2 Incidence of Federal Tax Filing and of Claiming the EITC

We next examine trends in the incidence of the filing of federal tax returns and of claiming the EITC. Over the course of the 1990s, the fraction of past welfare recipients that filed federal tax returns nearly doubled from 0.32 to 0.63 (Table 4.) This increase is roughly on par with the increase in rates of employment/labor market earnings displayed in Table 3. The most rapid increases in filing rates occurred in Los Angeles county (122.1 percent), followed by the Southern California (101.4 percent), Central Valley (100.6 percent) and Bay area (99.7 percent) regions. The lowest rates of increase occurred in the Central & Southern Farm (73.9 percent) and Northern and Mountain (57.9 percent) regions. By 1999, past welfare recipients had tax filing rates in excess of 60 percent for all regions of the State except the Northern and Mountain region counties.

Over the same period, the fraction of past welfare recipients in the State of California that claimed the EITC increased by 140.6 percent, going from 0.23 in 1992 to over half of sample (0.56) in 1999. As with tax filing rates, the increases in EITC claiming rates for adults previously on welfare were the highest in Los Angeles county (167.7 percent) and lowest for the Northern and Mountain (82.2 percent) and Central and Southern Farm (101.2 percent) regions.

4.3 EITC Participation Rates

In the previous subsection, we examined the incidence of households that claimed the EITC among *all* past welfare recipients. Some of the households in this group did not claim the EITC because they were not eligible for it.[10] As noted above, to be eligible for the credit, households had to have some earned income during a particular tax year and, putting aside for a moment the small EITC available to childless taxpayers, must have and care for at least one qualifying child. We do not have access to sufficiently detailed data about all sources of a household's earnings nor the number of children and their residency status that would allow us to determine a household's exact EITC eligibility status.

In what follows, we approximate EITC participation rates among those welfare households in California deemed eligible by counting the fraction of households in our sample with positive EDD wage and salary income in a given tax year that claimed the EITC on their federal tax returns. Most of the households who worked would be eligible for the EITC, since rarely will their earnings exceed the EITC eligibility limits (which in all years exceed $22,000) and most will have EITC-qualifying children (since they received welfare in the last quarter of the previous year). Moreover, similarly constructed participation rates have been used in past studies to assess

[9] Mean family income, in 2001 dollars, increased from $58,484 to $67,438 from 1992 to 1999 in the western states, which represents a 15.3 percent increase. See *Current Population Survey*, Annual Demographic Supplements.

[10] We also note that some households that claimed the credit may not have been eligible to receive it, i.e., their claims were noncompliant.

the rates of EITC participation in the U.S.[11] to which we can compare our estimates for California's welfare population.

Over the period 1992 through 1999, the EITC participation rate among those eligible for the credit increased by almost 50 percent, from 50 to 73 percent (Table 5). These participation rates roughly correspond with an estimated EITC participation rate of 75 percent for California in 1996 derived in a study funded by the Internal Revenue Service.[12] EITC participation rates varied by region of the State. In 1999 Los Angeles County had the highest rate (0.75) while the Bay Area counties had the lowest (0.67). Los Angeles County saw the highest rate of improvement in EITC participation rates over the 1990s, increasing by 50 percent, while the Northern and Mountain counties showing the slowest improvement in participation rates (27.9 percent).

Even with these increases in EITC participation, a sizeable percentage of California's past welfare recipients who are eligible for the EITC—27 percent statewide in 1999—were still not claiming, and thus not receiving, this credit by the end of the 1990s. Moreover, in some regions of the state, as much as one-third of those that may be eligible for the credit are not receiving it. It also appears that rates of EITC participation among California's welfare recipients is lower than for the nation as a whole. Estimates of the national EITC participation rates for households with children range from 80 to 86 percent,[13] with higher rates found for more recent years. That is, EITC participation rates for California's past welfare recipients are between 7 and 13 percentage points lower than those for the U.S. as a whole.

4.4 Average Amounts of EITC Received

Given the growth in generosity and coverage of the EITC over the 1990s, it is appropriate to ask how much of a financial benefit it provided to California's past welfare recipients. Among all households or assistance units during the 1990s, the average amount of the earned income credit received by the adults in past welfare assistance units increased more than 400 percent, going from an average of $243 per assistance unit in 1992 to $1,303 in 1999 (Table 6.) Among those past welfare units in which one or more adults earned income during the year, the average credit received more than doubled from $1,139 in 1992 to $2,345 in 1999. The average amounts of credit received and their rates of improvement over the 1990s varied by region of the State. For example, while the Central and Southern Farm and Northern and Mountain regions had among the highest average credits per assistance unit in 1992, $281 and $280, respectively, the

[11] See, for example, John Karl Scholz, "The Earned Income Tax Credit: Participation, Compliance, and Antipoverty Effectiveness," *National Tax Journal*, 41(1) (March 1994), pp. 63-87; Carolyn Hill, V. Joseph Hotz, Charles Mullin and John Karl Scholz, "EITC Eligibility, Participation and Compliance Rates for AFDC Households: Evidence from the California Caseload," final report for the State of California, University of California, Los Angeles, April 1999; and Len Burman and Deborah Kobes, "GAO Study of EITC Eligibility and Participation," Urban Institute, January 2001.

[12] Internal Revenue Service, "Participation in the Earned Income Tax Credit Program for Tax Year 1996," Washington, DC, January 31, 2002.)

[13] Scholz estimated EITC participation rates to range from 80 to 86 percent based on date from 1990 (see John Karl Scholz, "The Earned Income Tax Credit: Participation, Compliance, and Antipoverty Effectiveness," *National Tax Journal*, 41(1) (March 1994), pp. 63-87). The Internal Revenue Service estimated the national EITC participation rate for 1996 to be 82 and 87 percent, depending on data sources used (see Internal Revenue Service, 2002, "Participation in the Earned Income Tax Credit Program for Tax Year 1996," January 31, Washington, D.C.). Finally, the Government Accounting Office estimated that the EITC participation rate for households with children was 86 percent for tax year 1999 (see U.S. General Accounting Office, *Earned Income Tax Credit Eligibility and Participation*, GAO-02-290R, Washington, DC, December 14, 2001).

average amounts received by households in these regions by 1999 were below the State average.

4.5 Ratio of EITC Amount to EDD Earnings

In order to gain a sense of how important the EITC became as a source of income to households previously on welfare, we present estimates of the average ratio of the earned income credit received by households to their annual EDD earnings during the 1990s and by region of the State (Table 7). From 1992 to 1999, the importance of the EITC credit relative to EDD earnings more than doubled, going from 0.14 in 1992 to 0.34 in 1992. As with the credit amounts, the relative importance of the EITC to earnings varied by region of the state over the 1990s. As of 1999, for example, the highest ratios of credits to earnings were found in Los Angeles county, where the average household credit was 38 percent of EDD earnings, while, on average, the credits were only 28 percent of earnings of households residing in Bay Area counties.

5. Discussion

The trends and patterns presented above clearly indicate that EITC utilization rates and credits received by welfare recipient households in California increased over the 1990s and these credits provided increasing resources to these households, in both absolute and relative (to earnings) terms. Furthermore, our results suggest that an increasing fraction of households that appear to be eligible for the EITC actually claimed it over the 1990s.

Nonetheless, our estimates suggest that sizeable numbers of households who were eligible for the EITC did not claim it. While it is possible that our estimates of rates of EITC participation are incorrect given our relatively crude approximation to determining eligibility for the credit, our estimates of participation rates in California appear to be lower than comparably estimated rates for the U.S. as a whole. The fact that EITC participation is not 100 percent in California and that the State's participation rates are lower than the rest of the country suggest that a significant number of California's working poor are foregoing a sizeable amount of cash assistance for the federal government. Using numbers for the 1999 tax year,[14] our estimated EITC participation rate of 73 percent in California imply that as many as 860,000 households who were eligible for the EITC did not receive it. Even if California only achieved the participation rate for the nation as a whole (86 percent in 1999), our estimates suggest that some 480,000 households in California were not receiving tax credits for which they were eligible. Furthermore, based on the total credits paid out on 1999 tax returns in California, the State's working poor failed to receive between $800 million to $1.4 billion in credits.

The above estimated shortfalls from the EITC to California's working poor clearly suggest that value to understanding the factors accounting for California's low EITC participation rates and scope for actions that might improve EITC participation in the State. In the remainder of this section, we briefly discuss various factors that may have accounted for the increasing utilization of the EITC in California and for the lower-than-average EITC participation rates in the State.

With respect to factors that may have caused the increased rates of EITC claiming, we suggest several. First, over this period, the generosity and coverage of the EITC increased. As the result of federal legislation that took effect in 1994, the credit rate for filing units with two or

[14] For the 1999 tax year, 19.2 million households in the U.S. claimed the EITC on their tax returns for a total amount of $31.5 billion in credits. The corresponding numbers for the State of California were 2.3 million households and $3.9 billion in total credits.

more children and with adjusted gross incomes below approximately $9,000 were gradually increased from 19.5 percent in 1993 to 40 percent by 1996. The corresponding increases in "phase-in" tax credit rates for households with only one qualifying child were 18.5 percent in 1993 to 34.0 percent by 1996. Moreover, the size of the maximum credit was increased for households with two or more children (one qualifying child) from $1,511 ($1,434) in 1993 to $3,816 ($2,313) in 1999.[15] Finally, the coverage of the EITC was expanded so that by 1999 households with two or more (one) qualifying children received a credit if their adjusted gross incomes were below $30,095 ($26,928). In 1993, the corresponding adjusted gross income threshold for households with either one or two or more qualifying children was $23,050.[16] All of these changes in the EITC itself would lead to larger credits for those who qualified for it *and* increased the incentives of low-income households to work and file for the credit over the decade of the 1990s.[17]

Second, while the State of California experienced a severe recession beginning in 1991, over the remainder of the decade, California's economy improved, experiencing a substantial increase in jobs and growth in real wages.[18] This improvement in California's economy played a role in increasing the employment rates of the State's low-income population, increasing the fraction that had labor market earnings and, thus, were eligible to claim the EITC.

Third, the State's welfare system and its related welfare-to-work programs, instituted a variety of programmatic changes and rules aimed at moving those on public assistance "from welfare to work." These programs also likely increased the employment rates and the labor market earnings of past welfare recipients, which would have increased the fractions of low-income households, including past welfare recipients, who were eligible for the EITC.

With respect to the somewhat low EITC participation rates in California, especially in some of its counties, and the fact that the rates in California appear to be lower than those in the rest of the U.S., we offer several observations. First, we note that the differences in participation rates across regions and counties of the State may reflect differences in populations and their willingness to file tax returns. As the statistics in Table 2 make clear, the State's welfare populations differ in their racial and ethnic make-up by region. There is some evidence nationally that different racial and demographic groups have different propensities to file tax returns. For example, it has been argued that Hispanics and other ethnic groups in which there are sizeable proportions of recent, and possibly illegal, immigrants minimize their "contact" with government in California, especially since the passage of Proposition 189 in 1994. More careful analysis of the data is required in order to determine the importance of race and ethnicity in accounting for the observed regional differences in EITC participation rates.

Second, it is possible that households from different regions of the State may differ in their knowledge of the EITC and what they must do (file a tax return) to claim it. While the State, counties and various non-governmental organizations (NGOs) have undertaken efforts to in-

[15] These amounts are in current dollars.

[16] See V. Joseph Hotz and John Karl Scholz, "The Earned Income Tax Credit," in Robert Moffitt, ed., *Means Tested Transfers in the U.S.*, University of Chicago Press, forthcoming, for more details on the changes in the EITC over the 1990s.

[17] See Hotz and Scholz, *ibid*. for a summary of the existing evidence on the employment incentive effects attributable to the expansion of the EITC.

[18] See selected *California Statistical Abstract*, State of California Department of Finance, selected issues.

crease awareness of the EITC and, in some cases, provided tax preparation assistance for low-income populations, it is possible that these efforts were not equally distributed across the State. Furthermore, public service announcements and other information dissemination efforts about the EITC may have been constrained by differences in the coverage of radio and television media across the State. We note that the greatest gains in EITC participation and utilization over the 1990s occurred in Los Angeles county, which is a very large media market with print, radio and television outlets that specialize in and target various ethnic groups. Furthermore, it may be the case that commercial tax preparation services may be more readily available in Los Angeles County compared to other areas of the State, given the size and density of its population. Whether any or all of the above-noted factors play an important role in the incidence of EITC claims requires more refined and detailed analyses than can be undertaken in this report.

TABLE 1
Earned Income Tax Credit Parameters, 1992-1999 (in nominal dollars)

Year	Phase-in Rt %	Phase-in Range	Max Credit	Phase-out Rte (%)	Phase-out Range
1992[a]	17.6[1]	0-7,520	1,324	12.57	11,840 - 22,370
	18.4[2]		1,384	13.14	11,840 – 22,370
1993[a]	18.5[1]	0-7,750	1,434	13.21	12,200 - 23,050
	19.5[2]		1,511	13.93	12,200 – 23,050
1994	23.6[1]	0-7,750	2,038	15.98	11,000 - 23,755
	30.0[2]	0-8,245	2,528	17.68	11,000 - 25,296
	7.65[3]	0-4,000	306	7.65	5,000 - 9,000
1995	34.0[1]	0-6,160	2,094	15.98	11,290 - 24,396
	36.0[2]	0-8,640	3,110	20.22	11,290 - 26,673
	7.65[3]	0-4,100	314	7.65	5,130 - 9,230
1996	34.0[1]	0-6,330	2,152	15.98	11,610 - 25,078
	40.0[2]	0-8,890	3,556	21.06	11,610 - 28,495
	7.65[3]	0-4,220	323	7.65	5,280 - 9,500
1997	34.0[1]	0-6,500	2,210	15.98	11,930 - 25,750
	40.0[2]	0-9,140	3,656	21.06	11,930 - 29,290
	7.65[3]	0-4,340	332	7.65	5,430 - 9,770
1998	34.0[1]	0-6,680	2,271	15.98	12,260 - 26,473
	40.0[2]	0-9,390	3,756	21.06	12,260 - 30,095
	7.65[3]	0-4,460	341	7.65	5,570 - 10,030
1999	34.0[1]	0-6,800	2,312	15.98	12,460 - 26,928
	40.0[2]	0-9,540	3,816	21.06	12,460 - 30,580
	7.65[3]	0-4,530	347	7.65	5,670 - 10,200

Source: 1998 Green Book, Committee on Ways and Means, U.S. House of Representatives, U.S. Government Printing Office, page 867. 1998 through 2001 parameters come from Publication 596, Internal Revenue Service

[a] Basic credit only. Does not include supplemental young child or health insurance credits.

[1] Taxpayers with one qualifying child.

[2] Taxpayers with more than one qualifying child.

[3] Childless taxpayers.

TABLE 2
Demographic Characteristics of AFDC/TANF Assistance Units in California, 1992-1999

Tax Year	*Bay Area Counties*	*Southern Calif. Counties, excl. LA*	*Los Angeles County*	*Central & Southern Farm Counties*	*Northern & Mountain Counties*	*Central Valley Counties*
Proportion of Family Group (FG) Cases						
1992	0.781	0.756	0.739	0.718	0.747	0.771
1993	0.759	0.734	0.713	0.699	0.743	0.750
1994	0.751	0.721	0.710	0.692	0.744	0.743
1995	0.751	0.720	0.711	0.689	0.745	0.735
1996	0.751	0.724	0.733	0.667	0.733	0.733
1997	0.741	0.725	0.718	0.653	0.722	0.726
1998	0.738	0.743	0.720	0.660	0.711	0.722
1999	0.696	0.681	0.696	0.614	0.661	0.692
% Chg., 1992-99	*-10.9%*	*-9.9%*	*-5.8%*	*-14.5%*	*-11.5%*	*-10.2%*
Proportion of Family Group (UP) Cases						
1992	0.130	0.161	0.100	0.175	0.186	0.172
1993	0.134	0.168	0.111	0.185	0.194	0.184
1994	0.145	0.189	0.134	0.199	0.203	0.192
1995	0.147	0.195	0.142	0.205	0.201	0.198
1996	0.144	0.190	0.145	0.186	0.176	0.198
1997	0.141	0.186	0.140	0.175	0.160	0.202
1998	0.133	0.165	0.143	0.157	0.130	0.192
1999	0.123	0.183	0.131	0.149	0.117	0.187
% Chg., 1992-99	*-5.4%*	*13.7%*	*31.0%*	*-14.9%*	*-37.1%*	*8.7%*
Proportion of Cases with Mix of Aid Codes among Adults						
1992	0.089	0.083	0.160	0.107	0.067	0.057
1993	0.107	0.099	0.176	0.116	0.063	0.067
1994	0.103	0.090	0.156	0.109	0.053	0.065
1995	0.102	0.085	0.146	0.106	0.054	0.067
1996	0.105	0.086	0.122	0.147	0.090	0.069
1997	0.118	0.089	0.142	0.172	0.118	0.072
1998	0.129	0.092	0.137	0.183	0.159	0.086
1999	0.181	0.136	0.173	0.237	0.223	0.122
% Chg., 1992-99	*103.4%*	*63.9%*	*8.1%*	*121.5%*	*232.8%*	*114.0%*

Notes: *Sample:* Adults in AFDC/TANF assistance units in Quarter 4, Year *t*-1, derived from Medi-Cal Eligibility Determination System (MEDS) files. See Appendix for details.

Tax Year refers to the year in which any earnings were realized. In most cases, tax returns on these earnings are filed in the following calendar year, usually by April 15th.

Estimates for regions are averages of county-level statistics, weighted by county total population from the 2000 U.S. Census of Population.

TABLE 2 (Continued)
Demographic Characteristics of AFDC/TANF Assistance Units in California, 1992-1999

Tax Year	*Bay Area Counties*	*Southern Calif. Counties, excl. LA*	*Los Angeles County*	*Central & Southern Farm Counties*	*Northern & Mountain Counties*	*Central Valley Counties*
Average Number of Adult Females in Case						
1992	0.962	0.953	0.974	0.948	0.928	0.947
1993	0.961	0.952	0.972	0.941	0.922	0.940
1994	0.957	0.947	0.968	0.935	0.913	0.938
1995	0.957	0.944	0.964	0.929	0.913	0.938
1996	0.956	0.947	0.961	0.934	0.911	0.935
1997	0.959	0.948	0.966	0.936	0.912	0.938
1998	0.961	0.950	0.967	0.936	0.913	0.935
1999	0.961	0.949	0.972	0.938	0.903	0.937
% Chg., 1992-99	*-0.1%*	*-0.4%*	*-0.2%*	*-1.1%*	*-2.7%*	*-1.1%*
Average Number of Adults in Case						
1992	1.166	1.194	1.148	1.222	1.241	1.217
1993	1.181	1.206	1.156	1.227	1.244	1.226
1994	1.193	1.221	1.176	1.242	1.246	1.237
1995	1.196	1.222	1.175	1.246	1.247	1.247
1996	1.194	1.217	1.162	1.262	1.247	1.245
1997	1.196	1.213	1.172	1.263	1.257	1.256
1998	1.190	1.193	1.173	1.250	1.256	1.249
1999	1.195	1.205	1.183	1.255	1.263	1.264
% Chg., 1992-99	*2.5%*	*0.9%*	*3.0%*	*2.7%*	*1.8%*	*3.9%*
Average Number of Kids in Case						
1992	1.950	2.017	2.075	2.274	1.961	2.119
1993	1.961	2.035	2.098	2.272	1.960	2.115
1994	1.957	2.061	2.103	2.287	1.957	2.128
1995	1.960	2.061	2.100	2.274	1.954	2.137
1996	1.969	2.080	2.044	2.290	1.928	2.147
1997	1.995	2.106	2.069	2.318	1.947	2.196
1998	2.015	2.133	2.097	2.321	1.946	2.226
1999	2.053	2.167	2.132	2.337	1.945	2.253
% Chg., 1992-99	*5.3%*	*7.4%*	*2.7%*	*2.8%*	*-0.8%*	*6.3%*
Average Number of Kids, Ages 0-5, in Case						
1992	0.864	0.898	0.941	0.971	0.790	0.922
1993	0.865	0.908	0.946	0.974	0.799	0.915
1994	0.854	0.910	0.921	0.976	0.795	0.916
1995	0.833	0.882	0.881	0.948	0.778	0.889
1996	0.798	0.849	0.817	0.928	0.728	0.866
1997	0.773	0.823	0.798	0.906	0.711	0.850
1998	0.743	0.795	0.748	0.866	0.689	0.809
1999	0.727	0.771	0.728	0.856	0.682	0.783
% Chg., 1992-99	*-15.9%*	*-14.1%*	*-22.6%*	*-11.8%*	*-13.7%*	*-15.1%*

TABLE 2 (Continued)
Demographic Characteristics of AFDC/TANF Assistance Units in California, 1992-1999

Tax Year	*Bay Area Counties*	*Southern Calif. Counties, excl. LA*	*Los Angeles County*	*Central & Southern Farm Counties*	*Northern & Mountain Counties*	*Central Valley Counties*
Proportion of Cases with Some Whites						
1992	0.366	0.462	0.265	0.442	0.886	0.635
1993	0.351	0.433	0.259	0.423	0.875	0.629
1994	0.343	0.404	0.238	0.397	0.869	0.614
1995	0.338	0.389	0.229	0.387	0.868	0.605
1996	0.336	0.382	0.210	0.381	0.860	0.602
1997	0.325	0.371	0.206	0.369	0.856	0.592
1998	0.320	0.368	0.208	0.364	0.851	0.582
1999	0.303	0.348	0.207	0.352	0.845	0.572
% Chg., 1992-99	*-17.2%*	*-24.7%*	*-21.9%*	*-20.4%*	*-4.6%*	*-9.9%*
Proportion of Cases with Some Blacks						
1992	0.307	0.151	0.354	0.101	0.021	0.168
1993	0.299	0.149	0.327	0.098	0.023	0.169
1994	0.288	0.138	0.298	0.094	0.024	0.168
1995	0.285	0.137	0.288	0.090	0.023	0.170
1996	0.282	0.137	0.284	0.091	0.024	0.170
1997	0.285	0.142	0.285	0.094	0.027	0.176
1998	0.287	0.148	0.292	0.101	0.027	0.182
1999	0.293	0.155	0.300	0.104	0.026	0.187
% Chg., 1992-99	*-4.6%*	*2.6%*	*-15.3%*	*3.0%*	*23.8%*	*11.3%*
Proportion of Cases with Some Hispanics						
1992	0.235	0.337	0.403	0.414	0.058	0.153
1993	0.251	0.370	0.435	0.435	0.065	0.156
1994	0.272	0.411	0.481	0.470	0.070	0.176
1995	0.281	0.421	0.502	0.485	0.073	0.180
1996	0.282	0.422	0.514	0.497	0.076	0.178
1997	0.281	0.424	0.517	0.512	0.082	0.181
1998	0.274	0.420	0.509	0.512	0.082	0.181
1999	0.282	0.418	0.502	0.524	0.088	0.183
% Chg., 1992-99	*20.0%*	*24.0%*	*24.6%*	*26.6%*	*51.7%*	*19.6%*
Proportion of Cases with Some Other Race/Ethnicity (Primarily Asians)						
1992	0.220	0.190	0.154	0.201	0.151	0.177
1993	0.219	0.181	0.147	0.194	0.149	0.173
1994	0.213	0.169	0.139	0.175	0.142	0.168
1995	0.208	0.164	0.129	0.163	0.137	0.167
1996	0.209	0.163	0.117	0.151	0.134	0.167
1997	0.213	0.163	0.116	0.141	0.132	0.168
1998	0.221	0.162	0.115	0.136	0.136	0.171
1999	0.221	0.168	0.117	0.131	0.136	0.168
% Chg., 1992-99	*0.5%*	*-11.6%*	*-24.0%*	*-34.8%*	*-9.9%*	*-5.1%*

TABLE 3
Incidence of and Average Annual Wage and Salary Earnings for Adults on AFDC/TANF in California, 1992-1999

Tax Year	*Bay Area Counties*	*Southern Calif. Counties, excl. LA*	*Los Angeles County*	*Central & Southern Farm Counties*	*Northern & Mountain Counties*	*Central Valley Counties*	*Entire State*
Fraction of Welfare Recipients in Yr t-1:Qtr 4 Who Had Positive EDD Earnings in Tax Year t							
1992	0.39	0.37	0.31	0.41	0.42	0.39	0.37
1993	0.41	0.40	0.33	0.43	0.44	0.40	0.39
1994	0.47	0.45	0.41	0.49	0.48	0.46	0.45
1995	0.53	0.50	0.46	0.53	0.53	0.51	0.50
1996	0.58	0.54	0.51	0.57	0.55	0.54	0.54
1997	0.64	0.62	0.58	0.62	0.59	0.60	0.61
1998	0.69	0.67	0.64	0.66	0.63	0.66	0.66
1999	0.72	0.72	0.67	0.69	0.67	0.71	0.69
% Chg, 1992-99	*86.9%*	*94.1%*	*118.5%*	*67.6%*	*58.8%*	*84.6%*	*90.1%*
Average EDD Earnings in Tax Year t for Welfare Recipients in Yr t-1:Qtr 4							
1992	$2,519	$2,298	$1,946	$2,443	$2,582	$2,327	$2,273
1993	$2,578	$2,388	$2,098	$2,597	$2,552	$2,340	$2,373
1994	$3,047	$2,872	$2,645	$3,042	$2,994	$2,956	$2,872
1995	$3,914	$3,418	$3,138	$3,594	$3,618	$3,626	$3,461
1996	$4,457	$3,873	$3,589	$4,018	$3,837	$3,902	$3,893
1997	$5,425	$4,784	$4,537	$4,766	$4,472	$4,739	$4,774
1998	$6,330	$5,678	$6,141	$5,497	$5,114	$5,620	$5,858
1999	$6,771	$6,493	$6,101	$5,875	$5,691	$6,663	$6,271
% Chg, 1992-99	*168.8%*	*182.5%*	*213.5%*	*140.5%*	*120.4%*	*186.4%*	*175.8%*

Notes: *Sample:* Adults in AFDC/TANF assistance units in Quarter 4, Year *t*-1, derived from Medi-Cal Eligibility Determination System (MEDS) files. See Appendix for details.

Wage & Salary Data: Derived from match with data from Base Wage file, California Economic Development Department (EDD). See Appendix for details.

Tax Year refers to the year in which any earnings were realized. In most cases, tax returns on these earnings are filed in the following calendar year, usually by April 15th.

All dollar values are expressed in terms of 1999 dollars, using the CPI-U for the state of California.

Estimates for regions are averages of county-level statistics, weighted by county total population from the 2000 U.S. Census of Population.

TABLE 4
Rates of Filing Federal Tax Returns and Claiming the EITC for Adults on AFDC/TANF in California, 1992-1999

Tax Year	*Bay Area Counties*	*Southern Calif. Counties, excl. LA*	*Los Angeles County*	*Central & Southern Farm Counties*	*Northern & Mountain Counties*	*Central Valley Counties*	*Entire State*
Fraction of Welfare Recipients in Yr t-1:Qtr 4 Who Filed Tax Return in Tax Year t							
1992	0.31	0.32	0.29	0.35	0.37	0.32	0.32
1993	0.32	0.32	0.28	0.36	0.35	0.31	0.32
1994	0.36	0.38	0.36	0.40	0.39	0.36	0.37
1995	0.42	0.44	0.43	0.45	0.43	0.42	0.43
1996	0.47	0.48	0.48	0.49	0.46	0.45	0.48
1997	0.54	0.55	0.55	0.55	0.50	0.52	0.54
1998	0.58	0.59	0.59	0.58	0.53	0.57	0.58
1999	0.62	0.65	0.64	0.61	0.59	0.64	0.63
% Chg, 1992-99	*99.7%*	*101.4%*	*122.1%*	*73.9%*	*57.9%*	*100.6%*	*99.2%*
Fraction of Welfare Recipients in Yr t-1:Qtr 4 Who Claimed the EITC in Tax Year t							
1992	0.21	0.23	0.21	0.27	0.28	0.23	0.23
1993	0.23	0.25	0.21	0.29	0.28	0.23	0.24
1994	0.28	0.31	0.29	0.34	0.33	0.29	0.30
1995	0.34	0.37	0.36	0.39	0.37	0.35	0.37
1996	0.38	0.41	0.41	0.42	0.39	0.38	0.40
1997	0.44	0.48	0.49	0.49	0.44	0.45	0.47
1998	0.49	0.52	0.52	0.51	0.46	0.50	0.51
1999	0.53	0.57	0.57	0.55	0.51	0.55	0.56
% Chg, 1992-99	*149.6%*	*146.5%*	*167.7%*	*101.2%*	*82.2%*	*145.0%*	*140.6%*

Notes: *Sample:* Adults in AFDC/TANF assistance units in Quarter 4, Year *t*-1, derived from Medi-Cal Eligibility Determination System (MEDS) files. See Appendix for details.

Tax Return Statistics: Derived from matches of individuals from MEDS with Federal Personal Tax Returns and provided by the California Franchise Tax Board.

Wage & Salary Data: Derived from match with data from Base Wage file, California Economic Development Department (EDD). See Appendix for details.

Tax Year refers to the year in which any earnings were realized. In most cases, tax returns on these earnings are filed in the following calendar year, usually by April 15th.

Estimates for regions are averages of county-level statistics, weighted by county total population from the 2000 U.S. Census of Population.

TABLE 5
Estimates of EITC Participation Rates for Adults on AFDC/TANF in California, 1992-1999

Tax Year	*Bay Area Counties*	*Southern Calif. Counties, excl. LA*	*Los Angeles County*	*Central & Southern Farm Counties*	*Northern & Mountain Counties*	*Central Valley Counties*	*Entire State*
Fraction of Adults with Positive EDD Earnings in Tax Year t that Claimed the EITC							
1992	0.46	0.50	0.50	0.55	0.54	0.48	0.50
1993	0.46	0.51	0.48	0.56	0.54	0.48	0.50
1994	0.52	0.58	0.55	0.61	0.59	0.55	0.56
1995	0.56	0.63	0.63	0.66	0.62	0.59	0.62
1996	0.58	0.65	0.66	0.67	0.64	0.62	0.64
1997	0.63	0.70	0.70	0.71	0.68	0.66	0.69
1998	0.64	0.71	0.71	0.71	0.67	0.68	0.70
1999	0.67	0.73	0.75	0.73	0.69	0.71	0.73
% Chg, 1992-99	46.5%	44.4%	50.0%	33.3%	27.9%	46.6%	46.0%

Notes: *Sample:* Adults in AFDC/TANF assistance units in Quarter 4, Year *t*-1, derived from Medi-Cal Eligibility Determination System (MEDS) files. See Appendix for details.

Tax Return Statistics: Derived from matches of individuals from MEDS with Federal Personal Tax Returns and provided by the California Franchise Tax Board.

Wage & Salary Data: Derived from match with data from Base Wage file, California Economic Development Department (EDD). See Appendix for details.

Tax Year refers to the year in which any earnings were realized. In most cases, tax returns on these earnings are filed in the following calendar year, usually by April 15th.

Estimates for regions are averages of county-level statistics, weighted by county total population from the 2000 U.S. Census of Population.

TABLE 6
Average Amount of EITC Received by Adults on AFDC/TANF in California, 1992-1999

Tax Year	*Bay Area Counties*	*Southern Calif. Counties, excl. LA*	*Los Angeles County*	*Central & Southern Farm Counties*	*Northern & Mountain Counties*	*Central Valley Counties*	*Entire State*
Amount of EITC for Entire Sample in Tax Year t (i.e., was On-Aid in Yr t-1:Qtr 4)							
1992	$213	$238	$240	$281	$280	$227	$243
1993	$241	$275	$250	$323	$291	$249	$269
1994	$416	$488	$454	$545	$489	$436	$473
1995	$592	$714	$717	$765	$657	$642	$698
1996	$724	$846	$870	$892	$772	$751	$834
1997	$903	$1,075	$1,078	$1,072	$916	$943	$1,036
1998	$1,039	$1,222	$1,218	$1,185	$980	$1,075	$1,170
1999	$1,164	$1,370	$1,357	$1,277	$1,105	$1,245	$1,303
% Chg, 1992-99	*446.9%*	*475.6%*	*466.1%*	*355.2%*	*294.7%*	*448.9%*	*436.8%*
Amount of EITC, in 1999 Dollars, for Individuals in Filing Units that Claimed EITC in Tax Year t							
1992	$1,039	$1,101	$1,110	$1,085	$1,020	$1,061	$1,139
1993	$1,099	$1,186	$1,162	$1,178	$1,065	$1,124	$1,216
1994	$1,514	$1,688	$1,517	$1,657	$1,510	$1,551	$1,646
1995	$1,812	$2,031	$1,931	$2,018	$1,788	$1,879	$1,977
1996	$1,981	$2,202	$2,058	$2,188	$1,974	$2,015	$2,123
1997	$2,078	$2,323	$2,173	$2,285	$2,106	$2,145	$2,217
1998	$2,190	$2,453	$2,287	$2,389	$2,138	$2,186	$2,318
1999	$2,247	$2,478	$2,328	$2,400	$2,182	$2,250	$2,345
% Chg, 1992-99	*116.2%*	*125.1%*	*109.8%*	*121.1%*	*113.9%*	*112.1%*	*106.0%*

Notes: *Sample:* Adults in AFDC/TANF assistance units in Quarter 4, Year *t*-1, derived from Medi-Cal Eligibility Determination System (MEDS) files. See Appendix for details.

Tax Return Statistics: Derived from matches of individuals from MEDS with Federal Personal Tax Returns and provided by the California Franchise Tax Board.

Tax Year refers to the year in which any earnings were realized. In most cases, tax returns on these earnings are filed in the following calendar year, usually by April 15th.

All dollar values are expressed in terms of 1999 dollars, using the CPI-U for the state of California.

Estimates for regions are averages of county-level statistics, weighted by county total population from the 2000 U.S. Census of Population.

TABLE 7
Ratio of Average Amount of EITC Received to Average EDD Earnings for Adults on AFDC/TANF in California, 1992-1999

Tax Year	*Bay Area Counties*	*Southern Calif. Counties, excl. LA*	*Los Angeles County*	*Central & Southern Farm Counties*	*Northern & Mountain Counties*	*Central Valley Counties*	*Entire State*
Ratio of Average EITC Received in Tax Year t to Average EDD Earnings in Tax Year t							
1992	0.11	0.13	0.16	0.15	0.15	0.13	0.14
1993	0.12	0.15	0.15	0.16	0.15	0.13	0.14
1994	0.18	0.22	0.22	0.23	0.22	0.19	0.21
1995	0.23	0.29	0.34	0.29	0.26	0.26	0.29
1996	0.24	0.32	0.34	0.30	0.29	0.27	0.30
1997	0.25	0.32	0.36	0.31	0.31	0.28	0.32
1998	0.25	0.33	0.34	0.31	0.27	0.29	0.31
1999	0.28	0.33	0.38	0.32	0.29	0.30	0.34
% Chg, 1992-99	*150.7%*	*148.7%*	*141.8%*	*118.4%*	*97.6%*	*138.0%*	*142.8%*

Notes: *Sample:* Adults in AFDC/TANF assistance units in Quarter 4, Year *t*-1, derived from Medi-Cal Eligibility Determination System (MEDS) files. See Appendix for details.

Tax Return Statistics: Derived from matches of individuals from MEDS with Federal Personal Tax Returns and provided by the California Franchise Tax Board.

Wage & Salary Data: Derived from match with data from Base Wage file, California Economic Development Department (EDD). See Appendix for details.

Tax Year refers to the year in which any earnings were realized. In most cases, tax returns on these earnings are filed in the following calendar year, usually by April 15th.

All dollar values are expressed in terms of 1999 dollars, using the CPI-U for the state of California.

Estimates for regions are averages of county-level statistics, weighted by county total population from the 2000 U.S. Census of Population.

TABLE 8
Economic and Public Assistance Statistics for Regions of the State of California and the Entire State, 1999

Characteristic	*Bay Area Counties*	*Southern Calif. Counties, excl. LA*	*Los Angeles County*	*Central & Southern Farm Counties*	*Northern & Mountain Counties*	*Central Valley Counties*	*Entire State*
County Unemployment Rate	3.2	3.8	5.9	11.7	6.9	4.8	5.2
Median Household Income	$62,900	$49,820	$42,189	$38,323	$34,670	$45,195	$47,493
Median Family Income	$71,892	$55,564	$46,452	$42,456	$41,912	$52,611	$53,025
Per Capita Income	$30,770	$22,194	$20,683	$16,555	$18,429	$21,782	$22,711
% of All Individuals in Families with Income below Poverty Line	8.7	12.5	17.9	19.3	16.0	13.1	14.2
% of Children less than Age 18 in Fam. with Income below Poverty Line	10.2	16.2	24.2	25.5	20.5	17.3	19.0
% of Families with Income below Poverty Line	5.8	9.1	14.4	14.8	11.1	9.3	10.6
Ratio of AFDC/TANF FG Cases to Total No. of Females, Ages 15-44	0.095	0.144	0.242	0.309	0.289	0.223	0.189
Ratio of No. of Females on Medi-Cal to Total No. of Females, Ages 15-44	0.100	0.116	0.185	0.247	0.233	0.186	0.155

Notes: The data source for all entries, except for the unemployment, AFDC/TANF and Medi-Cal statistics, is the 2000 U.S. Census of Population. The unemployment rates are taken from the *California Statistical Abstract*, compiled by the State of California Department of Finance. The AFDC/TANF and Medi-Cal statistics are derived from MEDS data.

Estimates for regions are averages of county-level statistics, weighted by county total population from the 2000 U.S. Census of Population.

Figure 1. CDSS/DAPB Regions

APPENDIX:
Description of Samples Used and Data Used in Report

A1. Samples Used

The data used to produce the estimates in this report are for individuals and assistance units drawn from California's Medi-Cal Eligibility Data System (MEDS). All analyses are based on annual samples of adults who were on AFDC/TANF in the 4th quarter of Year *t*-1. For these samples, county-level averages were calculated for various variables as of Year *t*, which we refer to as *Tax Year t*. We refer to these annual samples as "*Past Welfare Recipient"* samples, given that they consist of individuals who were on welfare in Quarter 4 in the year prior to the Tax Year for which the outcome variables we analyze are recorded. *However, it is important to note that the adults in these annual samples may have been on aid (on AFDC or TANF) in Tax Year t; our strategy for constructing these annual samples ensures is that they were on welfare in the last quarter of the year prior to the Tax Year*.

The annual Past Welfare Recipient samples are, themselves, derived from an overall sample drawn from the MEDS population. The latter sample, referred to as the "*MEDS Sample*," was drawn as follows:

1. Individuals who were adults in the MEDS population are the unit of analysis in the MEDS Sample. These adults were at risk for being selected into the sample *only* in the first month in which they were on aid (AFDC-FG or AFDC-UP aid codes). The data are left censored in January 1987. Thus, in January 1987, a random cross-section of the data was selected. Then, in each month, a random sample of first-time entrants is added. This sampling scheme leaves the data representative of the MEDS population at each pointing time. Once an individual is selected into the sample, the individual remains in the sample throughout the rest of the sampling period.

2. The sampling probabilities used to draw the MEDS Sample varied by counties in the State. These sampling probabilities were chosen to obtain approximately equal sample sizes from each of the counties. However, in 36 of the 58 counties, the 100% sampling rate utilized still leaves the sample size smaller than that attained in the remaining 22 counties. Since individuals are at risk for sampling only in the first month observed, their sampling probability corresponds to the county in which they first receive aid. Realized sampling rates slightly differ from the sampling probabilities because a true random sample was drawn.[19] Sampling weights were constructed.

3. Three criteria we used for selecting individuals and cases for inclusion in the MEDS Sample:[20]

[19] The RAND Corporation determined the sampling probabilities for each county and drew the sample from the MEDS. RAND is in the process of producing a document detailing how the sample was constructed.

[20] While not reported herein, we reproduced all of the estimates presented in Tables 1 through 4 and Table 6 including some or all of the cases that were excluded based on these three criteria. None of trends described in this report were sensitive to relaxing any of the above criteria.

- We included/excluded cases from our MEDS Sample based on the quality of the information available for the case. In particular we included the following types of cases in our analysis sample: (a) Family Group (FG) cases that contained only one adult—most typically a female—and those Unemployed Parent (UP) cases that contained exactly two adults and (b) FG and UP cases that contained extra adults who were *not* on aid in the fourth quarter of the previous year. We excluded those cases from our analysis for which there were too many adults or which had conflicting aid codes. For example, one FG adult and 2 UP adults in the same case or cases that had five adults were excluded. The latter cases were excluded because we lack confidence in our ability to construct households that correspond to tax filing units in such cases.[21]

- Cases were also included/excluded cases based on having valid Social Security numbers (SSNs), including only those cases from the MEDS that had SSNs that had been verified with the Social Security Administration (SSA). Excluding cases without valid SSNs was necessary, given our need to match adults in the MEDS assistance units to EDD wage and salary data and federal tax returns. Eliminating unverified SSNs and cases with too many adults (described above) produces our MEDS Sample.

- Based on the wage and salary data from EDD Base Wage Files (described below) on individuals, we excluded cases that had reported wage income from more than 20 different employers in a given year.[22]

A2. Sample Weights

We construct two distinct sets of weights for the observations in the MEDS Sample (and, thus, the annual Past Welfare Recipient samples). The first set is individual-based weights intended to make the sample representative of the MEDS populations when performing analysis at the individual level. These weights are the inverse of individuals' sampling probabilities. In particular, let p_i be the sampling probability of the *i*th individual. Then, $1/p_i$ is the *i*th individual's weight.

The second set is case-based weights. These latter weights make the sample representative of the MEDS population when performing analysis at the case level. They are the inverse of the probability that the case was sampled. For a case with *n* individuals, the following equation gives the probability the case is sampled:

$$p^c = 1 - \prod_{i=1}^{n} (1 - p_i) .$$

Therefore, the appropriate weight for case-level analysis is $1/p^c$. These weights adjust for both the over-sampling of cases from small counties and the over-sampling of cases containing many

[21] We believe that most of the excluded cases represent multiple households (assistance units) that fall under a common serial number in the MEDS database. However, counties differ in their use of assistant unit codes and MEDS lacks documentation on these codes. Therefore, we are unable to disaggregate assistance units within a serial number.

[22] For example, some people have 200+ employers over a calendar year. In these cases, it is likely that multiple people are using the same SSN.

members.

A2. Data Sources Used, Matching Procedures, and Variables Analyzed

Demographic Characteristics of Assistance Units of Sampled Individuals:

For individuals and their associated AFDC/TANF assistance units in the annual Past Welfare Recipient samples, data on the demographic characteristics of these assistance units are drawn from the MEDS database. Summary statistics for these characteristics are presented in Table 2, by year and by region of the State.

Wage and Salary Earnings Variables:

To obtain wage and salary earnings data for the sampled adults, and other adults in the sampling unit at the time of inclusion in the annual Past Welfare Recipient samples, quarterly wage and salary earnings data from the California Employment Development Department (EDD) Base Wage Files were matched for these individuals by Social Security numbers (SSNs). We constructed *case-level earnings* for the annual Past Welfare Recipient samples based on the individual-level earnings recorded in the EDD Base Wage files for all of the adults in the assistance unit of those individuals on aid in Quarter 4, Year *t*-1. These case-level earnings were used to determine whether anyone in the associated assistance unit had positive EDD earnings. Table 3 contains estimates of both the incidence of positive EDD earnings and average levels of earnings, in 1999 dollars, by year and region of the State.

Federal Tax Return Information, including Incidence of EITC Claiming and Credit Amounts:

To obtain information on the rates of tax return filing, claiming of the EITC and amounts of the credit received adults in our samples, the California Franchise Tax Board (FTB) matched information from federal tax returns to the individuals in the assistance units in our annual Past Welfare Recipient samples based on SSNs and performed all of the analyses involving tax return data. *All tax return for individuals in these cases were aggregated into a case- or assistance-unit-level return. If anyone in an assistance unit filed a tax return or claimed the EITC, then the case was recorded as filing a return and/or claiming the EITC. Finally, the size of the credit received by the case is the sum of the credits received by individuals in the case (joint returns, common in UP cases, were not double counted).* The FTB provided only summary statistics from analyses of these individual- or assistance-unit-level data to the authors for use in this Report.

EITC Eligibility, Participation and Compliance Rates for AFDC Households: Evidence from the California Caseload

Carolyn J. Hill
University of Chicago
cjhil@cicero.spc.uchicago.edu

V. Joseph Hotz
UCLA
hotz@ucla.edu

Charles H. Mullin
Vanderbilt University
charles.mullin@vanderbilt.edu

John Karl Scholz
University of Wisconsin-Madison
jkscholz@facstaff.wisc.edu

Prepared for

State of California
Gray Davis, Governor
Grantland Johnson, Secretary, California Health and Human Services Agency
Rita Saenz, Director, California Department of Social Services
Gerald H. Goldberg, Executive Officer, California Franchise Tax Board

April 1999

Acknowledgements

The analysis presented in this report was conducted under contract number 95ASPE270A, granted by the Assistant Secretary for Planning and Evaluation of the U.S. Department of Health and Human Services (DHHS) and by the state of California Department of Social Services (CDSS). The authors wish to thank Werner Schink and Leslie Raderman of the California Department of Social Services and George Ramsey of the California Franchise Tax Board for their support and assistance in completing this project. The authors also wish to thank Diane Gibson-Cubbin and Jeffrey Liebman for helpful comments on an earlier draft of this report. The conclusions drawn in this report are those of the authors and do not necessarily reflect those of DHHS or CDSS.

Table of Contents

Executive Summary

In this report, we examine the eligibility for and participation in the earned income tax credit (EITC) for low-income families in California during the 1990s. The EITC is a federal income tax credit available to working poor families, with the amount of credit determined by the number of eligible children and by family income. The EITC has become the largest cash or near-cash program available to low-income families in the U.S., making it one of the cornerstones of the nation's antipoverty effort. The effectiveness of the EITC depends, in part, on whether those who are eligible for the credit actually receive it and whether it is actually claimed.

We examine EITC eligibility, participation and non-compliance rates for a special group of potential EITC beneficiaries, namely low-income households in California who received AFDC during the 1990s. Our sample consists of AFDC assistance units from four counties—Alameda, Los Angeles, San Bernardino and San Joaquin—that were part of the California Work Pays Demonstration Project (CWPDP), a waiver demonstration to assess the impacts of a series of reforms to the AFDC program in the state. For this sample we have data from administrative records from county welfare offices on AFDC participation, AFDC payments and the income reported to county welfare departments for the purpose of determining AFDC payments. These data were supplemented with quarterly wage and salary information from the state's unemployment insurance records as well as taxable income, by source, and EITC credits from federal income tax returns for the years 1993 and 1994. The latter information was obtained by matching state unemployment insurance wage earnings records and IRS tax returns data for the adults in the assistance units included in the demonstration project.

EITC Eligibility Rates: Using several alternative operational measures of the criteria for EITC eligibility, we estimate that between 21 and 53 percent of CWPDP assistance units were eligible for an earned income tax credit in 1993 and 1994 (Table 3a). The rates of eligibility were higher among AFDC-U units (i.e., primarily units headed by two adults) than AFDC-FG units (i.e., units headed by a single adult), and the rates are higher among assistance units that have only recently gone on welfare compared to samples that include long-term welfare recipients.

EITC Participation Rates: With respect to EITC participation rates, which are determined using data from federal tax returns, we find that between 42 and 84 percent of eligible assistance units filed a tax return and, as such, claimed the EITC in 1993 or 1994 (Table 3a). While these rates are lower than those estimated for all EITC eligible households in the U.S.—the latter estimated participation rates are around 80 percent—they are not substantially lower, especially among AFDC-U assistance units in the CWPDP. As with eligibility rates, EITC participation is higher among AFDC-U assistance units than AFDC-FG units and higher among recent entrants to welfare versus long-term recipients.

EITC Non-Compliance Rates: Finally, we estimate that the proportion of assistance units in our samples that filed tax returns and claimed the EITC who we estimate were not eligible for the credit—which we call the EITC non-compliance rate—ranged from 10 to 57 percent depending upon which operational definition of eligibility was used (Table 3a). We find that rates of non-compliance were higher for AFDC-FG units compared to AFDC-U cases and that non-compliance was more prevalent among those units who had only recently entered the AFDC rolls. We

note that our estimated rates of non-compliance for a population of households taken from welfare rolls are not substantially different from recent estimates of rates of non-compliance obtained by the IRS from a nation-wide audit of the EITC program in 1988. (The latter audit study found that approximately one-third of all claimants of the EITC were not eligible for the credit.) At the same time, we must add a cautionary note concerning our findings on non-compliance. As explained below, our classifications of EITC eligibility are based on income measures that *understate* the earnings of households. Furthermore, the measures we use to determine whether households meet the presence of qualifying children are less than ideal. Accordingly, our estimates of non-compliance rates must be treated with caution, as they are likely to be subject to bias.

While most of the assistance units in our sample were still on welfare during the years (1993 and 1994) in which we analyze their EITC status, not all are. Between 1 and 10 percent of AFDC-FG units and between 1 and 13 percent of AFDC-U households were no longer on AFDC in the years we analyze EITC eligibility, participation and compliance (Tables 4a and 4b). But, the estimated EITC eligibility, participation and non-compliance rates for those units that were on AFDC for at least one month in a particular tax year are not substantially different from those for all units in the CWPDP (Table 3b). As such, our findings indicate that a non-trivial proportion of households on AFDC—as well as those who were recently on welfare—work enough to qualify for an earned income tax credit and, of those who qualify, a substantial fraction actually received assistance from this program.

It is important to note that the data we use to analyze EITC eligibility, participation and compliance for households from the California AFDC caseload are subject to potential problems with accurately measuring the two key determinants of EITC eligibility: the annual earned income of adults in the household and the extent to which these adults have responsibility for the care of children during a particular tax year. With respect to income, our measures are limited to what households reported to county welfare departments and wage and salary earnings for which unemployment insurance taxes were paid. We do not have separate measures of self-employment or casual labor income on individuals beyond what those who file tax returns report to the IRS. With respect to measuring the "presence-of-children" requirement for EITC eligibility, we only have information on this status in months when the household was on AFDC. In order to assess the extent to which these potential measurement problems may bias our findings, we also investigated the comparability of our alternative measures of income and the presence-of-children across available data sources.

With respect to children for which adults may be responsible, we compared the reported number of children in each AFDC assistance unit with the number of exemptions claimed on tax returns for the assistance units that filed federal tax returns (Tables 5a and 5b). We find that between 57 and 80 percent of these households claimed the same number of children under the AFDC program and on their tax returns, with the lower match rates occurring among AFDC-FG units. When there are discrepancies in reporting, the number of exemptions tends to be *less* than the number of children reported in the AFDC assistance unit. While this evidence is less than conclusive, it suggests that the limitations of our information on the presence-of-children are not likely to impart sizable biases on our estimates of EITC eligibility, participation and compliance. To the extent there are biases, it is likely that our estimates of EITC eligibility and participation

rates *understate* and our estimates of non-compliance *overstate* the respective true rates in our sample.

Several conclusions emerge from our analysis of differences in measures of income across AFDC, unemployment insurance and IRS records. First, for between 75 and 82 percent of the assistance units in our sample, there is little difference in the levels of labor earnings they report to county welfare departments and those for wages and salary covered by unemployment insurance (Tables 7a and 7b). These relatively high rates of agreement result from the fact that the majority of households in our sample report no income to the AFDC program *and* had no earnings subject to unemployment insurance taxes. For households that have *positive* earnings that are subject to UI, we find that between 55 and 69 percent of these units *underreport* their earnings to the welfare system. Almost all of this underreporting is due to households with wage earners reporting that they had none to the AFDC program.

Second, we tend to find substantial agreement between income reported to the IRS and earnings subject to UI. These two measures are approximately the same for between 64 and 83 percent of the households in our sample (Tables 8a through 9b). However, among those adults who filed tax returns, both the Adjusted Gross Income (AGI) and Line 7 Wage and Salary Earnings on these returns are more likely to *exceed*, rather than be less than, earnings subject to UI. Given that both the AGI and Line 7 income categories are more inclusive measures of household income than are wages subject to UI, this pattern is not surprising. For example, the reported AGI and/or Line 7, Wage and Salary earnings would include self-employment income or income from casual labor.

Because self-employment income could be overreported by households in order to qualify for an earned income tax credit, we also investigate the patterns of its reporting. In general, we find that only a small fraction of adults in our sample claimed any self-employment income on their tax returns (Tables 4a-4c). Between 4 and 6 percent of households on AFDC-FG claimed to have income from self-employment and the corresponding rate was approximately 11 percent for those on AFDC-U. Moreover, among those adults we classified as eligible for the EITC, the incidence of self-employment is even lower. However, we do find that among households that we classified as ineligible for the EITC, the incidence of self-employment income reported is much higher. Between 7 to 11 percent of AFDC-FG cases and 25 to 26 percent of AFDC-U cases reported having some self-employment income on their returns. The operational definitions of EITC eligibility used in our analysis are based on income measures that either are found to be underreported (i.e., reports to the AFDC program) or do not include self-employment income. These biases in our income measures provide another reason to suspect that our estimates of EITC eligibility and participation are *biased downward* and that our non-compliance rates are *biased upward* for the low-income population on welfare that we have analyzed.

There appear to be several conclusions one can draw, albeit tentatively, from the findings of this study about the role of the EITC and with the incentives that programs like AFDC and the EITC create for reporting and behavior. First, we find that a non-trivial proportion of households on AFDC were eligible for and claimed the earned income tax credit during the first half of the 1990s. In particular, we find that between 12 to 20 percent of the households in our sample on AFDC-FG and between 20 and 33 percent of those on the AFDC-U program actually received a earned income tax credit for years in which they were also receiving AFDC. While not huge, this

degree of overlap between the utilization of these two programs is not negligible. Furthermore, while the average annual credit received by households in our sample is small—between $166 to $381 per year for those on AFDC-FG and between $236 to $615 for those on AFDC-U—this amount increased from 1993 to 1994, amounting to between 5 to 12 percent of the average annual benefits these households received from the AFDC program. In short, this study documents an important, and non-negligible, overlap between the AFDC and EITC programs.

Our evidence on the EITC "involvement" of those on welfare also provides some encouragement that the provisions of the recently passed Personal Responsibility and Work Opportunities Act which seek to encourage greater work among this population may succeed. These findings are somewhat surprising for at least two reasons. First, California experienced a severe recession during this period; one would have expected that the employment opportunities of low-income adults would have been severely limited. Second, most previous studies have found that the rates of labor force participation among those on AFDC are very low. The basis for the latter conclusion is the incidence of labor market earnings reported to welfare departments by AFDC recipients. Our study documents, as have others, that these reports severely underreport such income. Moreover, we find that this underreporting is not simply attributable to the failure to report "off-the-books" earnings from casual labor, as suggested in a recent ethnographic study of welfare recipients by Kathryn Edin. In fact, we find that earnings from employment that is verifiable and readily documented are substantially underreported as well.

1. Introduction

The earned income tax credit (EITC) is a federal income tax credit available to working poor families. First adopted in 1975, the EITC was promoted as a way to relieve the burden of Social Security payroll taxes on low-wage working parents. The original EITC equaled 10 percent of earnings (which was also the combined employer and employee share of payroll taxes) up to a maximum credit of $400 for taxpayers with at least one child. In real terms, the EITC was roughly constant between 1975 and 1990, but the maximum credit has nearly tripled since 1990. In fiscal year 1998, the EITC is expected to cost the federal government $27.8 billion, making it the largest cash or near-cash program available to low-income families. As such, it has become a cornerstone of the nation's antipoverty efforts.

A family's EITC is determined both by the number of eligible children and by a family's earnings from work. There are three distinct income ranges relevant to the EITC: the phase-in, flat, and phase-out ranges (Figure 1). For taxpayers with one child in 1997, the phase-in range for family income is from zero to $6,500 and the credit is 34 percent of income. The flat range is from $6,500 to $11,930 and the credit is $2,210 (34 percent of $6,500). The phase-out range is income above $11,930 and the credit is the maximum of zero and $2,210 minus 15.98 cents for every dollar of income above $11,930. Families with two or more children are entitled to a larger credit ($3,656, or 40 percent of $9,140).[1] Table 1 summarizes EITC parameters for recent years, including those covered in our data. Unlike most credits and deductions in the federal individual income tax system, the EITC is refundable: if the amount of the credit exceeds what the family owes, the household receives a payment from the U.S. Treasury for the difference.

The effectiveness of the EITC in supporting the working poor depends, in part, on whether those who are eligible for the credit actually receive it. A family receives the EITC by filing a tax return; however, some families that do not have positive tax liabilities are not required to file tax returns. To the extent that low-income families, such as those who are on AFDC, do not file, the antipoverty effectiveness of the EITC is reduced, since families with incomes considerably less than the legal filing thresholds may be entitled to substantial EITC benefits.[2] Given the broad-based support for policies, such as the Personal Responsibility and Work Opportunities Act of 1996, which encourage the transition from welfare to work, it is of interest to examine the extent to which households that received cash assistance under the now defunct AFDC program were eligible for and claimed the EITC. Previous studies of welfare or the EITC provide little information about the proportion of "welfare" populations that qualify for or claim the EITC. For these reasons, the first goal of this study is to provide new, reliable information on the fraction of AFDC recipients that were eligible for and received the EITC during the early

[1] The credit for families with two or more children is phased out at a rate of 21.06 percent, starting at an income of $11,930.

[2] For the 1990 tax year, Scholz (1994) found that the EITC participation rates for the eligible population was between 80 and 86 percent. (Scholz (1994) used data from the Survey of Income and Program Participation, matched by Social Security number to individual income tax returns to calculate EITC eligibility and participation. The variation in Scholz's estimates of participation rates arises from different assumptions made in the calculation.) We note that the EITC participation rates are high compared to other public assistance programs. For example, the AFDC participation rate is estimated to be between 62 and 72 percent among those who are eligible for this program and the participation rate for the Food Stamp program is between 54 and 66 percent (Blank and Ruggles, 1993).

1990s.[3]

The second goal is to investigate non-compliance with EITC eligibility requirements.[4] Such non-compliance can take many forms. As noted above, a family's eligibility for the EITC depends on their earned income and the number of children for which they are responsible. Non-compliance can be with either of these criteria. Some examples include: (a) the parents of a child who file separate tax returns, each claiming the same child as a dependent; (b) taxpayers who claim non-existent children, neighbors' children, or nonresident nieces or nephews; and (c) taxpayers who misreport their earned income. Some of these cases may not reflect the intent to commit fraud, while others do. A recent study of the EITC for Tax Year 1994 conducted by the Internal Revenue Service found that $4.4 billion, or 25.8 percent of total EITC claims, exceeded the amount to which taxpayers were eligible. Evidence of such rates of non-compliance has lead House Ways and Means Chairman Bill Archer to conclude that the credit is "one of the most abused programs on the books."[5]

The IRS study of EITC non-compliance, and the previous ones conducted by the IRS, does not estimate rates of EITC non-compliance for filers who qualified for cash assistance programs, such as AFDC. Nothing is known about whether failure to comply with EITC eligibility requirements is higher or lower among households who received, or had recently received, AFDC. In this study, we focus on the rates of non-compliance among households on, or recently on, AFDC in California that claimed the credit. We emphasize that our estimates of noncompliance may be subject to biases arising from the income and presence-of-children measures used to estimate EITC eligibility.

In this study, we examine the incidence of EITC non-compliance, as well as of EITC eligibility and participation, using information from several alternative data sources. As described below, we make use of information from the administrative records of households on AFDC in four counties in California. These data are linked with information on the earnings for adults in these households from the state's unemployment insurance wage and salary system. In addition, we obtained information from the federal tax returns, including reported earnings and EITC claims, for those in our sample who filed tax returns. (The tax return information was obtained under an interagency agreement between the California Department of Social Services and the State of California Franchise Tax Board, the state's taxing authority.) We use such data to construct several alternative measures of EITC eligibility and participation for households we classify as being eligible and not eligible for an EITC.

Verifying whether or not families meet either of the two eligibility requirements for claiming the EITC with respect to earned income and dependent children is difficult, even when thorough audits of tax returns are performed. As we discuss below, limitations of the information

[3] Between the time the data for this study were collected and the research was conducted, the Aid to Families with Dependent Children (AFDC) program was eliminated and replaced by Temporary Assistance for Needy Families (TANF). Throughout the report the phrase "AFDC recipient" denotes members of an assistance unit under the former AFDC program.

[4] Throughout this study, we use the term "noncompliance" to refer to erroneous EITC claims caused by negligence, mistakes, confusion and fraud.

[5] *Wall Street Journal*, April 30, 1997, p. 1.

available to us from administrative records constrain our ability to determine a household's true EITC eligibility status and, thus, whether they are in compliance. To assess the robustness of our findings on eligibility and noncompliance, we examine the consistency, across the administrative data sources available to us, of the measures of earned income and the presence of dependent children and earned income. Specifically, we compare family size and number of adults recorded in AFDC case files with the number of dependents and adults claimed on tax returns. We also look across several data sources—self-reports of earned income to public assistance and to tax authorities, as well as employer-reported earnings recorded in the unemployment and disability (UI/DI) insurance system—to determine whether income is reported consistently to different authorities. Our objective is to ascertain the extent to which our measures of EITC non-compliance, as well as eligibility, are likely to be biased and the likely nature of these biases.

We investigate, in some detail, the extent to which the households in our sample report earnings from self-employment when filing tax returns for two reasons. The first concerns the potential biases in our estimates of EITC eligibility and non-compliance. To determine whether a sample household meets the EITC eligibility criteria, we are limited to using either household income reported to county welfare offices for purposes of benefit determination or annual wages and salaries for which employers pay unemployment insurance taxes. The latter measure of income does not include income from self-employment or exempted types of employment and, as will be discussed below, the former is only measured when households are on the AFDC rolls and, even then, is likely to understate the total earnings of adult members of the household, especially self-employment and casual labor income. Because self-employment income can be used to qualify for an EITC, our methods for determining EITC eligibility may understate the true rates and, as such, overstate EITC non-compliance. In order to get a sense of the magnitude of this bias, we examine the incidence of self-employment income reported on federal tax returns by adults in our sample households.

The second reason for examining the reporting of this type of income concerns the incentives the structure of the EITC program creates for misreporting this source of income. Adults may fraudulently claim higher levels of self-employment income so as to qualify for a higher credit. In a recent study of EITC reporting for the tax year 1994, the General Accounting Office found that the percentage of EITC recipients in the Phase-In range of the credit reported more profits from self-employment than did those taxpayers in either the Phase-Out and Flat ranges (13.7 percent versus 11.8 percent).[6] More generally, those who filed for the EITC had a higher incidence of self-employment income than did other individual taxpayers. In 1994, 15.2 percent of all EITC recipients reported self-employment income on Schedule C of their tax returns; this accounted for 7.3 percent of the total income reported by recipients.[7] In contrast, only 10.4 percent of all individual taxpayers reported any self-employment income in 1993 and it accounted for only 4.4 percent of the income they reported.[8] The higher rates of reporting of this source of income by low-income filers, relative to all individual filers, may reflect, in part, involvement of the former individuals in more "casual" forms of private contract work. However, it may also reflect fraudulent reporting, especially given the inherent difficulty of verifying the accuracy of the

[6] See GAO Briefing Report, *Earned Income Credit: Profile of Tax Year 1994 Credit Recipients*, June 1996, p. 38.

[7] See GAO Briefing Report, *Earned Income Credit: Profile of Tax Year 1994 Credit Recipients*, June 1996, p. 30.

[8] See GAO Report, *Tax Administration: Tax Compliance of Nonwage Earners*, August 1996, Figure 2, p 7.

reporting of this type of income.

The remainder of the report is organized as follows. Section 2 describes the data sources used in our analysis. In Section 3, we report on our findings for the related issues of EITC eligibility and participation for AFDC recipients in California. In section 4, we consider the coverage and comparability of income reports and presence of dependent children across the administrative data sources available to us for the families in our sample. The final section provides a brief summary and draws conclusions from our research.

2. The Data

All of the work described in this report exploits a unique dataset constructed through cooperative agreements with several agencies of the California state government and with UC-Data, a research data repository at the University of California-Berkeley. The core sample is drawn from four California counties that participated in the California Work Pays Demonstration Project (CWPDP), an evaluation of waivers granted by the federal government to the state's AFDC program. This demonstration project is overseen by the California Department of Social Services (CDSS), the agency that controls access to the CWPDP data. Data on the adults in the CWPDP households, or assistance units, were matched to administrative files to obtain quarterly wages and salary earned during 1993 and 1994 in jobs covered by unemployment and state disability insurance programs for all adults in these households; access to the latter data was obtained by CDSS under an interagency agreement with the California Employment Development Department (EDD) which maintains this information. We shall refer to this data set as the "UI/DI" or "EDD" wage or earnings data. A match for adults in the CWPDP was also performed with federal tax returns data for the tax years 1993 and 1994; access to this latter data was provided under a strict set of protocols negotiated by CDSS and the California Franchise Tax Board (FTB) to insure the confidentiality of the tax returns information of individuals.[9] We refer to the information from these matched tax returns as the "IRS" data. The linking was based on matching Social Security numbers of the CWPDP adults with those for EDD and IRS administrative records. Finally, while the resulting linked data provides a rich and unique set of data on income and tax credits, we have only a limited amount of demographic information for the household, which was drawn from AFDC and food stamp records.

In the following subsections, we provide more detailed descriptions of the sampling plan and structure of the data used in our analysis, and describe the data obtained from the UI/DI systems and the IRS databases.

2.1 The Baseline Sample for the EITC Project: Assistance Units in the California Work Pays Demonstration Project (*CWPDP*)

Beginning in 1992, California obtained a series of waivers from the U.S. Department of Health and Human Services to enact a number of changes to its AFDC program.[10] The impact of

[9] Under the arrangement, all linking and analysis of data involving or containing tax returns information was performed by FTB staff on FTB computers. Summary statistics from these analyses were all that was provided to the authors of this report.

[10] The following changes were enacted:

these waivers on program participation and costs to the federal government were evaluated under the CWPDP, which consisted of a sample of assistance units that were on AFDC at the time of their entry into the demonstration and that resided in one of four counties within the state: Alameda, Los Angeles, San Bernardino and San Joaquin. To monitor the impact of the waivers on these assistance units, information was extracted from the public assistance administrative records systems in these counties. The data from this source are referred to as the "CWAD" data, referring to the County Welfare Administrative Database. The CWAD contains monthly, case-level information on: (a) the number of adults and children in an assistance unit, (b) the unit's reported Gross Earned Income (GEI) used to determine their monthly AFDC payment[11] (except in San Joaquin); (c) the unit's GEI for purposes of determining their food stamp benefits;[12] and (d) their monthly AFDC payments. We focus our analysis of EITC eligibility, participation, and non-compliance rates for the population of welfare recipients in California from which our sample is drawn.

Eliminate the 100 hour (per month) work limitation for remaining eligible for AFDC-U: This change does not affect the eligibility for the AFDC Unemployed Parent program (AFDC-U), but does affect conditions under which one remains eligible. This change affected only AFDC-U cases; the rule was not in effect for AFDC-FG cases.

Remove time limit for $30 and 1/3 income disregard: This change became effective in July 1993. Prior law required that if AFDC recipients earned income after four months on AFDC, they were subject to a 100 percent benefit reduction rate (BRR). The new law removed the 4-month time limit and allowed eligible AFDC recipients to keep $30 plus one-third of their earnings.

Reduce Maximum Aid Payment: The maximum amount of AFDC cash aid was reduced by a total of 15 percent.

Implement of Cal-Learn program: This program encouraged pregnant teens and teen parents to stay in or return to school by providing child care, transportation, and other assistance, and by creating disincentives for bad grades or for dropping out of school.

Increase personal resource limits and allowance of savings accounts for education: This provision raised the limits on personal resources and automobile stock that AFDC recipients can hold and remain eligible for AFDC. It allows recipients to retain up to $5,000 per family in a restricted account to be used for a child's post secondary education, for down payment on a home, or for starting a business. These new rules do not apply to resources allowed at the time of eligibility determination, in which case the old rules still apply.

The California Alternative Assistance Program (CAAP): This provision enables AFDC-eligible persons to decline an AFDC cash grant, but still receive Medi-Cal (the Medicaid program in California) and child care assistance.

Change employment services programs: Several provisions were implemented to make California's Job Opportunities and Basic Skills (JOBS) training program and the Greater Avenues for Independence (GAIN) program more work-oriented.

[11] AFDC gross earned income is defined as the sum of gross income for all individuals in the assistance unit, where gross income for each individual is calculated as the sum of total earnings from wages and salaries, in-kind earned income, and net self-employment income (gross self-employment income less business expenses). Earned income is income received in cash or in kind as wages, salary, commissions or profit from activities such as businesses or farming, in which the recipient is engaged as a self-employed individual or as an employee, or income received as disability benefits. Exempt earned income *not* included in the AFDC computation are earnings of a child derived from participation in JTPA programs, all earnings of a child under 19 years of age who is either a full-time student or a half-time student not employed full-time, income from College Work Study Program, and the first $50 of any child or spousal support paid to the assistance unit.

[12] Food stamp gross earned income includes all wages and salaries of an employee, gross income from a self-employment enterprise excluding the cost of doing business, training allowances, payments to VISTA volunteers, JTPA earnings (unless earned by dependent under 19), and strikers' benefits.

To evaluate the impact of the provisions of California's AFDC program relative to its program prior to the enactments of the waivers, assistance units (or cases) included in the CWPDP were randomly assigned to either a control or treatment status. Control group cases received AFDC under the rules in place as of September 1992. The treatment group, along with AFDC recipients in the rest of the counties of California, was subject to the provisions granted under the waivers. Since the reforms were designed to increase the attractiveness of work among welfare recipients, one might expect to see differences in EITC eligibility between treatment and control cases. While this is a potentially interesting topic to examine, separate analyses of EITC eligibility, participation and compliance for treatment and controls groups are not presented in this report. Analysis of the impact of the waivers on EITC eligibility and participation is a potential subject for future investigation.

The four research counties for the CWPDP have distinct geography, populations, welfare caseloads, and welfare departments. Los Angeles and Alameda Counties contain major urban centers, and San Bernardino and San Joaquin counties are their neighboring rural areas. San Joaquin is located in the agricultural region known as the Central Valley, while San Bernardino is part of Southern California's desert region and is the largest in terms of land mass of the four counties. Los Angeles County has the largest population of any California county with nearly nine million residents in 1990. San Bernardino and Alameda Counties each have 1.2 to 1.5 million residents and San Joaquin has approximately 500,000 residents. Not surprisingly, Los Angeles County has the largest AFDC caseload, with more than 285,000 cases in 1992; the other three counties have caseloads in the 30,000-60,000 range with San Joaquin's caseload being the smallest. However, San Joaquin County had the highest percentage of its population on AFDC of all four counties in 1990. In addition, San Joaquin has a relatively high percentage of its caseload in the Unemployed Parent component of the AFDC program (20 percent) compared to the other three county caseloads (10 to 15 percent).

For purposes of assessing rates of EITC eligibility, participation and compliance with these data, we excluded some of the assistance units in the full CWPDP sample. In particular, we imposed the following exclusions to obtain our *analysis sample*:[13]

1. *Exclusion of all "late entering" adults from assistance units in our analysis sample:* Recall that data in the CWAD are maintained for members of the case at the time of sampling as well as for members who enter the case at a later time (for both Original and Replenishment Sample cases). As part of the CWPDP, these late entrants into a case were not submitted for a match to Base Wage File records and as such we cannot observe earnings information for them. Hence, we exclude these persons from our analysis.

2. *Exclusion of cases from analysis sample that were from San Joaquin County and members of the CWPDP treatment group:* Administrative problems in San Joaquin resulted in incomplete and inconsistent information for these cases.[14] We do use data on control group households

[13] When constructing a measure of the income of assistance units, we did not include any earnings of children, i.e., assistance unit members under the age of 18. Children are likely ineligible for the EITC; even if a child had earnings, that child is not eligible to claim the EITC if she is a qualifying child for EITC purposes of another person.

[14] See Becerra, *et al.* (1996) for a complete discussion of the problems with data for treatment group cases in San Joaquin county.

that reside in San Joaquin county.

3. *Exclusion of all "child-only" cases from our analysis sample:* We excluded all cases that were "child-only" cases that were eligible for AFDC.[15] In this type of assistance unit, the adults who live with the children are ineligible for AFDC, because the adults are undocumented workers or have been sanctioned out of the case for violating aid regulations. We eliminated these cases from our analysis because we do not have Base Wage File earnings or tax data for the adults in the household. Child only cases are quite prevalent in California; as such, this exclusion accounted for 23 to 39 percent of the full CWPDP sample.

4. *Exclusion of assistance units with three or more adult members:* The EITC statuses of these "complex" households are potentially interesting ones to study, since the scope for manipulating the reporting of family structure to qualify for the EITC are significantly higher for this group than for one- or two-parent only households, Nonetheless, we dropped them from our analysis sample because of inadequate information about relationships between these adults and the children in the unit that we will discuss below. We note that at the time of sampling into the CWPDP, up to 5 percent of the AFDC-U cases, but less than 1 percent of the AFDC-FG cases contained three or more adults.

Background characteristics for the assistance units in the resulting CWPDP Analysis Sample are provided in Tables 2a and 2b. (In the Appendix to this report, we compare the characteristics of these assistance units with those in the full CWPDP sample.) In the remainder of this report, the unit of analysis is the *case* or *assistance unit*. We also use sample weights in all estimates presented below, where the weights are constructed to reflect oversampling of certain types of AFDC cases (AFDC-U versus AFDC-FG cases) and the caseload that existed at the time a case was drawn into the sample.

An important feature of our analysis sample, as well as the full CWPDP sample, was the way assistance units from the caseloads in the four counties were drawn into the CWPDP. The CWPDP is comprised of two somewhat distinct samples. One sample of assistance units was drawn at the inception of the CWP Demonstration Project in October 1992; we shall refer to this sample as the "Original Sample." While representative of the caseload at a point-in-time, studies by Bane and Ellwood (1983), Ellwood (1986), and O'Neill, Bassi, Wolfe and Hannan (1984) have made clear that such samples are disproportionately made up of assistance units that are "welfare dependent." The second sample in the CWPDP consists of a random sample of assistance units that entered the AFDC caseload in 1993. We refer to this sample as the "Replenishment Sample."[16] Previous studies have found that a significant proportion of new entrants remain

[15] We define a child as being 18 years old or younger throughout the year. This age limit corresponds to the age of a qualifying child for EITC purposes. The range noted in the text arises from differences across our subsamples. Thirty-nine percent of the AFDC-FG cases in the 1993 Replenishment Sample are child-only cases while 23 percent of the AFDC-U cases in the 1992 sample were child-only.

[16] The Original Sample was randomly drawn from cases on aid in October 1992. The Replenishment Sample consists of assistance units that entered the AFDC caseload over the months from March to December 1993. Replenishment cases selected in March 1993 could not have received assistance in California in October or November 1992. Cases new to aid in December 1992, January 1993, or February 1993 were eligible to be drawn for the March 1993 sample. Replenishment cases selected in April 1993 through November 1993 could not have received assistance between December 1992 and the sampling month. Beginning in December 1993, a case was eligible for inclu-

on welfare for only a relatively short period of time.[17] Furthermore, a study of welfare dynamics[18] indicates that most new entrants exit from AFDC to employment. Thus, it is more likely that the typical assistance unit in the Replenishment Sample will be eligible for and claim the EITC in the years following their entry to the CWPDP than will a typical unit in the Original Sample. Because of these past findings concerning patterns of labor force participation rates of AFDC cases by their time-on-welfare, all of our analyses are conducted separately for these two samples throughout this Report.

Another important feature of our CWPDP analysis sample concerns the type of AFDC assistance units were receiving at the time they were drawn into the CWPDP. In particular, our analysis sample consists of two types of assistance units: (1) needy one-parent households with children who qualify for the basic AFDC program, which are referred to as "AFDC-FG" or "family group" cases, and (2) two-parent households who are needy because the principal earner has been unemployed and qualifies for the Unemployed Parents program and are referred to as the "AFDC-U" assistance units. In the analysis that follows, we report all results separately for AFDC-FG and AFDC-U units in our data for several reasons. First, in contrast to the eligibility for the AFDC-FG program, the principal earner in a household must have worked in the past in order to be eligible for the AFDC-U program.[19] Second, assistance units in the latter program usually have two adults present, whereas most households on the AFDC-FG program contain only one adult, the parent of the children in the unit. Given these differences, one would expect, *a priori*, to find differences in the incidence of work-related income across these two types of households and, thus, in their likelihood of being eligible for or participating in the EITC. Accordingly, we present separate estimates for AFDC-FG and AFDC-U units in all of the analyses that follow.

2.2 Matched Data from the Employee Development Department (EDD) Base Wage File

The state of California's Employment Development Department (EDD) maintains a Base Wage File containing quarterly wage and salary earnings of individuals working in jobs that are subject to the state's unemployment insurance (UI) and disability insurance[20] (DI) programs.[21] We were provided with wage data from EDD files for all of the adults in CWPDP households at

---

sion in the Replenishment Sample if it had not received assistance at any time during the previous twelve months. New cases initiated in November 1992 are not represented in the initial or Replenishment Samples.

[17] For example, Bane and Ellwood (1983) estimate that 65 percent of new entrants leave the caseload in two years or fewer.

[18] See Gritz and MaCurdy (1992).

[19] To qualify for the AFDC-U program, the principal earner must: (a) have 6 or more quarters of work in any 13-calendar-quarter period ending within 1 year prior to application for assistance; or (b) have received or been eligible to receive unemployment compensation within 1 year prior to application for assistance. A quarter of work is a quarter in which an individual earns at least $50 or in which the individual participated in an authorized job training program.

[20] Disability Insurance is a state-mandated insurance plan in California that is financed through payroll deductions. Disability Insurance provides benefits to eligible workers who suffer loss of wages when they are unable to work because of non-work related illness, injury or pregnancy. Most California workers are covered by this program.

[21] The file generally includes individuals paid cash wages of more than $100 in a calendar quarter and domestic workers paid cash wages more than $750 in a calendar quarter.

the time they were sampled by UC-DATA. The matching was based on matches of Social Security numbers.

An employee's record in the Base Wage File is likely to be an accurate source of earnings information if the position is covered by UI/DI. The information in this file is provided by employers whose workers are covered by these social insurance programs. It would appear that they have little incentive to underreport their employees' earnings because the wage and salary base on which UI taxes are calculated is quite low and the employer does not pay additional UI taxes once an employee's earnings exceed this base. Furthermore, a self-policing mechanism ensures accuracy of employers' earnings reports, as they must be consistent with the employee's report if he or she applies for benefits under either the UI or DI program.

At the same time, the state's UI and DI systems do not cover all employment. The earnings of certain types of workers are exempted from UI/DI coverage, such as those for self-employed workers who do not elect UI/DI coverage; federal employees (military or civilian), nonprofit organization employees, railroad employees, students working for a school, college, or university; and casual labor paid less than $50 in a calendar quarter and working fewer than 24 days in that or the preceding quarter. As such, this measure, while easily verifiable and accurate, does not include all of the sources of income for which an earned income tax credit can be claimed.

2.3 Matched Data from Federal Income Tax Returns

The state of California's Franchise Tax Board, the state's taxing authority, attempted to match federal tax returns for all adult members of the assistance units in our CWPDP analysis sample. Again, this matching process used Social Security numbers as the match criteria.[22] Based on the results of this matching process, we determined which families (and its members) filed a federal tax return. For those that did, we obtained the following information for both the 1993 and 1994 tax years: (a) the filing status of filer (i.e., married, single, head-of-household); (b) the number of exemptions claimed by the filer; (c) the filer's Adjusted Gross Income (AGI), total Wage and Salary Income (reported on Line 7 of the IRS 1040 Form); (d) amounts of income by source (e.g., wages and salary, self-employment, etc.); and (e) whether the EITC was claimed and the amount of the credit granted.

2.4 The Economic Environment over the Study Period

The period over which we examine EITC eligibility and participation in California, 1993 and 1994, was far from typical. During the first half of the decade of the 1990s, the state of California economy experienced a fairly sharp recession, as did the rest of the nation. The growth rate of employment in all industries in 1991 was -2.69 percent compared to 2.65 percent in 1990. (See Figure 2.) Moreover, the severity of this recession varied across regions of the state and the four counties we examine in this study. For example, total employment declined by 5.68 percent in Los Angeles county in 1991, while, in this same year, total employment grew slightly in San

[22] The processing of all data involving tax returns data was conducted by the California Franchise Tax Board on their computers; under the agreement with the Board, neither the analysis team for this project nor employees of the California Department of Social Services were given direct access to any tax return data.

Bernardino county (0.86 percent). By 1993 and 1994, the California economy was recovering the 1991 recession, with employment *declining* at an annual rate of 0.52 percent in 1993 and *growing* at a rate of 1.26 percent in 1994. Again, while the recovery was occurring in all four counties used in our analysis over the period of our analysis, the strength of the recovery varied across them. While we do not attempt to systematically adjust our EITC eligibility and participation estimates, one must be cautious in generalizing from our estimates to other stages of a business cycle.

3. EITC Eligibility, Participation, and Non-Compliance

In this section, we examine the EITC eligibility, participation and non-compliance rates for our Analysis Sample of households on AFDC in California during the early 1990s. Examining the EITC eligibility and participation of this population is of interest, given the desire for policies to encourage the transition from welfare to work. Based on official national statistics on the share of the AFDC caseload that reported any earned income, it would appear that only a relatively small fraction of the AFDC caseload would qualify for an EITC. In 1994, 8.9 percent of all AFDC cases reported any earned income, which is 30 percent lower than the rate in 1980, with an average of $394 per month.[23] Among those households on the basic AFDC program covering single-parent households with children, the corresponding rate was somewhat lower, with 8.4 percent of such cases reporting some earned income which averaged $374 a month; among two-parent households on the AFDC-U program, the corresponding percentage was much higher (25.9 percent) with an average monthly amount of $471. Given that almost all AFDC households have children, these figures on earned income while on welfare would indicate very low rates of eligibility for EITC among those on welfare.

The above rates are based on what AFDC households report to the AFDC program for purposes of determining their monthly benefits. Several recent studies suggest that the actual percentage of AFDC recipients with earned income—and, thus, potentially eligible for an EITC—are much higher. Based on data from several waves of the Survey of Income and Program Participation (SIPP) during the 1980s, Spalter-Roth, *et al*. (1995) find that as many as 43 percent of female-headed households on AFDC households had worked for pay. In a study of single AFDC mothers in four sites (Chicago, Boston, San Antonio and Charleston, S.C.), Edin (1995) found that 46 percent of her sample earned income, which they did not report to AFDC authorities, where the vast majority of these cases (over 90 percent) either were working for cash or under a false identity in the informal economy.

While it is unclear whether the latter earnings would be reported to the IRS to claim an EITC, this evidence strongly suggests that the rates of *eligibility* for a credit may be much higher than suggested by official AFDC statistics. We now turn to the crucial question of how to measure the elements of the EITC eligibility criteria, namely, a household's annual income and presence of qualifying children.

[23] See *The Green Book* (1996), Table 8-28. Nationally, the percentage of AFDC households reporting any earned income has declined by almost 30 percent since the beginning of the 1980s, although these rates have grown somewhat since 1985.

3.1 Operational Definitions of EITC Eligibility Criteria

In Figure 3, we provide a schematic representation of the various "cells" in which we wish to classify assistance units with respect to their EITC eligibility and participation. The columns of Figure 3 split the sample into households eligible for and ineligible for the EITC, while the rows indicate a household's status with respect to filing a tax return and claiming the EITC. Our first objective is to determine EITC eligibility, i.e., the percentage of our sample who were in Cells I, II or III during a particular tax year. We then turn to determining the fraction of those eligible that participated, i.e., the ratio of assistance units in Cell I to those in Cells I, II or III.

As discussed in the Introduction, a household eligible for the EITC must have a qualifying child and income in a specified range.[24] With respect to qualifying children, they must: (1) be under age 19 at the end of the calendar year; (2) reside with the adult filer for more than half the year;[25] and (3) satisfy a relationship test.[26] With respect to income, one or more of the parents (or adults) in the household must have earned some income during the year to qualify for a credit.

While these criteria are relatively straightforward, their measurement is less so: limitations of our data complicate our ability to determine the EITC eligibility status of assistance units in our analysis sample. With respect to determining whether a household met the presence-of-qualifying-children requirement, the only source of information we have on children is that contained in the CWAD, and this information is less than ideal. This is because information about the family structure—including the number and ages of children—is only available for the months subsequent to entering the CWPDP in which the household is receiving AFDC benefits.[27] In the other months, we do not know for sure which, if any, children remain with the adult heading the unit. In short, the CWAD data do not allow us to unambiguously determine whether a household meets the presence-of-a-qualifying-child requirement for EITC eligibility. With respect to measuring income, recall that the two measures of income that are available to us are monthly GEI reports to the county welfare department for determining AFDC benefits and EDD wage and salary earnings, i.e., earnings from employment subject to UI and DI in California. Unfortunately, both of these measures are less than ideal for use in determining a household's EITC eligibility. As the evidence from the Spalter, *et al*. and Edin studies suggest, reports of GEI may understate a household's earnings. As we have already noted, earnings from EDD records do not include earnings from self-employment, even though such earnings can qualify for a credit. Fur-

[24] Filers with no children could claim the EITC in 1994. These persons are unlikely to be present in our dataset and are not relevant for our analysis.

[25] Persons under age 24 who are students, and persons any age who are permanently and totally disabled are also qualifying children for EITC purposes if they live with the filer for more than half the year. Our data do not provide information on these types of persons, however, they are unlikely to be an important part of our sample. Additionally, foster children are eligible for EITC purposes if they live with the filer for the entire year. Our data contain information about foster children, however, they represent a very small portion of our sample and, because of data limitations, we do not impose the requirement that they live with the foster parent for a full year.

[26] To satisfy the relationship requirement, the qualifying child must be a son, daughter, adopted child, grandchild, stepchild, or foster child of the taxpayer. Due to data limitations, we cannot identify relationships of household members and are unable to test the relationship requirement.

[27] We suspect that the CWAD information on children present for these months is fairly accurate, since this information is necessary to determine AFDC benefits.

thermore, low human capital workers may be in jobs that are not covered by the UI insurance system.

We deal with these measurement limitations in several ways. We investigate the consistency of these measures with information on exemptions claimed and earnings reported on tax returns for those assistance units that do file tax returns. The results of this latter investigation are presented in Section 5 below. We also construct several alternative operational criteria for classifying households as to their EITC eligibility status and use these to form corresponding estimates of EITC eligibility rates, as well as EITC participation and non-compliance rates. By examining the sensitivity of these estimated rates to the criteria used, we attempt to assess, in a limited way, the robustness of our conclusions about the importance of the EITC for welfare populations.

More precisely, we use the following *six* ways of combining information on income and presence-of-children to define a household's *EITC Eligibility*:

<u>*Criterion A*</u>*:* A household is eligible for EITC in a particular tax year if

(a) EDD Wage Earnings > 0 *and*
(b) it is assumed to meet qualifying child criteria in the tax year by virtue of having at least one child at the time it entered the CWPDP.

<u>*Criterion B*</u>*:* A household is eligible for EITC in a particular tax year if

(a) Max{GEI for AFDC; GEI for Food Stamps; EDD Wages} > 0 *and*
(b) it is assumed to meet qualifying child criteria in the tax year by virtue of having at least one child at the time it entered the CWPDP.

<u>*Criterion C*</u>*:* A household is eligible for EITC in a particular tax year if

(a) EDD Wage Earnings > 0 *and*
(b) it was on AFDC (and had at least one child in the AFDC assistance unit) for at least *one month* in that year.

<u>*Criterion D*</u>*:* A household is eligible for EITC in a particular tax year if

(a) Max{GEI for AFDC; GEI for Food Stamps; EDD Wages} > 0 *and*
(b) it was on AFDC (and had at least one child in the AFDC assistance unit) for at least *one month* in that year.

<u>*Criterion E*</u>*:* A household is eligible for EITC in a particular tax year if

(a) EDD Wage Earnings > 0 *and*
(b) it was on AFDC (and had at least one child in the AFDC assistance unit) for at least *six months* in that year.

<u>*Criterion F*</u>*:* A household is eligible for EITC in a particular tax year if

(a) Max{GEI for AFDC; GEI for Food Stamps; EDD Wages} > 0 *and*
(b) it was on AFDC (and had at least one child in the AFDC assistance unit) for at least *six months* in that year.

These measures differ in the stringency of the criteria used for income and presence of children. Some of these definitions, Criteria *A, C,* and *E,* have a *stricter* standard for meeting the income component of eligibility, requiring that at least one adult in the household had earnings subject to UI/DI, whereas the other measures, Criteria *B, D,* and *F,* allow a household to meet this component of eligibility if either EDD or earnings reported to the welfare department are positive. With respect to meeting the qualifying child component, the criteria vary with respect to the amount of "verified" evidence of a child being present during the tax year. We note that in this dimension, being able to verify that a child is present also "requires" that the household had to have spent an increasing proportion of the year on the AFDC caseload. Looking across these six criteria, it would appear that using *Criterion B* to define EITC eligibility will produce an *upper bound* for EITC eligibility rates as it the least stringent with respect to meeting both the income and presence-of-qualifying-children components of the eligibility requirements, while using *Criterion E* should produce a *lower bound* for these rates as it imposes the most restrictive standards for both components. Finally, we consider use of *Criterion A* for determining EITC eligibility to be our *intermediate* estimates of rates given that they are based on an income measure which we know to be accurately measured and on a measure of children that does not impose the link between subsequent AFDC receipt and eligibility that are inherent in the measures requiring verifiable evidence for the presence of children.

3.2 Estimates of Rates of Eligibility for the EITC

In Tables 3a and 3b, we summarize the range of estimates of EITC eligibility rates for the various samples, tax years, and types of assistance units for our CWPDP sample. Table 3a contains estimates for all assistance units in our Analysis Sample, regardless of whether the unit received any AFDC benefits in the particular tax year analyzed. Table 3b provides estimates for the subsamples of assistance units from our Analysis Sample which received AFDC benefits for at least one month during the particular tax year that is analyzed. Thus the eligibility (as well as participation and compliance) rates presented in Table 3b are estimates for a population of *current* AFDC recipients, while those in Table 3a are for a population of households which *have recently been* on AFDC. In each of these tables, we provide lower and upper bound estimates of rates of EITC eligibility as well as those based on our intermediate criterion (Criterion A).

Several conclusions emerge from an examination of the estimates of eligibility rates in these two tables. First, regardless of the criteria used, we estimate that sizable fractions of the households either on or recently on AFDC would be eligible for the EITC. Even the lowest rate we estimate—for AFDC-FG, regardless of their AFDC receipt status in the tax year—indicates that 21.5 percent of the units appear to be eligible for a credit. Using Criterion A, the criterion most easily defended with our available data, we estimate that EITC eligibility rates range from between 26 to 43 percent of assistance units. Note that these rates are based solely on finding EDD earnings for adults in these households, since the criterion for children, in essence, plays no role. As such, these rates indicate that substantially higher percentages of AFDC households in California have earned income compared to what would be predicted based on estimates of rates of labor market earnings reported by AFDC recipients nationally. (Recall that those rates indicated that only 8.4 percent of single-parent AFDC recipients reported any income to welfare departments in 1994.) In fact, based on our upper bound estimates, these rates may be as much as 57 percent of AFDC-U recipients. In short, our estimates suggest that a non-trivial share of households on, or recently on, AFDC have verifiable earnings and, as such, appear to be eligible

for receiving some assistance under the EITC program. This evidence has implications for the potential role the EITC program has for mitigating some of the impacts of the reductions in financial assistance under TANF.

Second, we find that larger shares of assistance units on AFDC-U are eligible to receive an EITC than is the case for households under AFDC-FG program. Using Criterion A, we find that between 30 and 43 percent of AFDC-U households may be eligible for an EITC, with the rate possibly being as high as 57 percent. It is not surprising that eligibility rates for AFDC-U households are generally higher than rates for AFDC-FG households since at least one adult in an AFDC-U household is likely to have earnings at some point in the year.

Third, we find that lower eligibility rates for households in our Original Sample compared to those in our Replenishment sample. (The higher estimates for each group are seen in 1994 tax year eligibility for households in the Replenishment Sample.) Recall that longer-term AFDC recipients who typically do not work dominate the Original Sample.

Fourth, a comparison of the eligibility rates in Table 3b with those in Table 3a indicate that the above conclusions about EITC eligibility apply to households who *were on AFDC* for some time during the various tax years. As such, these findings indicate that a sizable share of our households are in line to participate in both the AFDC and EITC programs and suggest that gaining a better understanding of the interactions between the two, and their consequences of encouraging work and reducing dependence on cash assistance, is worth exploring. We discuss the actual extent of joint participation in these two programs when we discuss EITC participation below.

Before concluding our discussion of EITC eligibility rates, we briefly consider the likely ways in which our estimates are biased. As discussed earlier, it is likely that eligibility rate estimates are biased *downward* because we do not have precise information about self-employment income or Social Security numbers used to match program administrative data with tax returns. For example, we may misclassify a self-employed person as ineligible for the EITC because the Base Wage File does not include earnings of self-employed persons.[28] We have independent information from the tax return data on the presence of self-employment income. As we shall discuss later in this report, the incidence of reporting self-employment income reported to tax authorities among this population is very low, although we do find slightly higher rates of self-employment income for those in this population that we classify as ineligible. We might also classify a tax-filing unit as ineligible for the EITC due to lack of dependents in the program administrative data. This person could rightfully claim the EITC, yet we would classify her as an ineligible participant. However, the range of assumptions used for presence of children has to provide an upper bound for eligibility rates, since, under two of the eligibility criteria we use, we assume all units have a qualifying child for the EITC. Hence, it does not appear that this source of misclassification is likely to impart much downward bias in our estimates of eligibility rates.

[28] Our estimates of EITC eligibility may be biased downward biased because we do not find a match with EDD unemployment insurance data. We discuss the coverage of EDD data in the report's section on coverage and compatibility across datasets.

3.3 Estimates of EITC Participation Rates and Amount of Credits Received

Based on our alternative operational criteria for determining EITC eligibility, we examine the proportion of assistance units who had an adult who claimed the EITC on their tax returns. More precisely, the EITC participation rate is defined to be the number of eligible filers who claim the credit as a fraction of all eligible persons in the population, *whether or not they file a tax return*.[29] Estimates of these rates are also found in Tables 3a and 3b.[30]

Several broad conclusions emerge from an examination of our estimates of rates of EITC participation for households in the CWPDP. First, we find that between 42 and 84 percent of eligible households in our Analysis Sample actually claimed a credit (see Tables 3a and 3b). These rates vary by the criteria used to determine eligibility, by the type of AFDC assistance and by sample (Original versus Replenishment). (As can be seen from the percentages of all households who filed tax returns given in Tables 3a and 3b, failure of "eligible" households to participate generally results from their failure to file a tax return.) Second, as with rates of eligibility, we find systematic differences in participation rates by type of assistance, sample and tax year. Participation rates were generally higher for AFDC-U than for AFDC-FG households, ranging from 57 to 84 percent for AFDC-U households and from 42 to 54 percent for AFDC-FG households. We also find that EITC participation is higher for households in our Replenishment samples as compared to those in our Original samples. It is also the case that EITC participation rates for AFDC-FG households in the Original Sample increased slightly over the two years, from around 47 to 50 percent. However, participation rates for AFDC-U households in the Original Sample fell slightly over the two years, from 67 to 65 percent (Table 3a). Finally, we find little difference in the rates of EITC participation, associated with our various eligibility criteria, across our full Analysis Sample (Table 3a) and those households in this sample that were on AFDC for at least part of the tax year analyzed (Table 3b).

The estimated participation rates for this welfare population presented in Tables 3a and 3b are lower than previous estimates for the entire population. For example, Scholz (1994) found that between 80 and 86 percent of all those individuals in the U.S. that were eligible for the EITC participated in the program during the latter part of the 1980s. Furthermore, the failure of households in our welfare-based sample, who appear to be eligible for a credit, to file for it, substantially lowers the share of the sample that actually receives assistance from this program. Using our upper and lower bounds, as well as intermediate, estimates of EITC eligibility and participation rates, our results imply that only between 10 and 24 percent of households on, or recently on, the AFDC-FG program in California actually claimed a credit; the corresponding rates for

[29] We calculate this rate by dividing Cell I by Cells I+II+III in Figure 3.

[30] We also examined a slightly different way of calculating the EITC participation rate, which reflects the possible taxpayer response to an IRS notification policy during 1993 and 1994. Specifically, the IRS does not automatically compute and award the credit to eligible nonparticipants (Cell II in Figure 3). Instead, a notice is sent informing the filers they might be eligible for the EITC if they file amended returns. Thus, two estimates of the participation rate are needed to reflect the IRS policy for notifying possible EITC recipients. The first measure (the number of eligible claimants divided by all eligible persons in the population) assumes that none of the eligible nonparticipants submitted amended returns in response to the IRS notice. The second measure assumes that all eligible nonparticipants respond to the IRS notification, file amended returns, and receive the credit. We focus in this report on the first measure of participation, but acknowledge that it is the lower bound of a participation rate. The rates based on this second method do not differ much from those presented in Tables 3a and 3b.

households on the AFDC-U program are almost twice as high, ranging from 16 to 46 percent.

There are several possible reasons why apparently-eligible AFDC recipients did not file tax returns and claim an EITC. It may be that our estimates of the rates of eligibility are wrong and many of these households do not file because they are not eligible; presumably, this was because they did not meet the presence-of-a-qualifying-child criteria, since we use a verifiable measure of income—wage earnings from EDD—to determine the income component of eligibility.

Alternatively, it may be that the size of the credit to which they are entitled is not sufficient to induce them to file a return. Most of the households in our sample have earnings in the Phase-In range of the EITC. In this range, the lower a household's earnings the lower the credit they can claim. We record, in Tables 4a, 4b and 4c, a number of characteristics about the earnings and AFDC receipt of our various samples.[31] Among the households we classify as eligible for the EITC, average EDD earnings are consistently lower for households who do not file tax returns. Average annual EDD wage earnings for these households range from $1,893 to $4,417, while EDD earnings for EITC-eligible households that file tax returns range from $4,484 to $9,680. Furthermore, in 1993 we find that the vast majority of EITC claimants from our Original Sample that we classified as being eligible for the credit in our sample of current, or recent, AFDC recipients had Adjusted Gross Incomes (AGI) which put them in the Phase-In range of the EITC program (see column (i) in Table 4a). In 1994, we find that a slightly lower fraction of eligible claimants had income in the Phase-In range, with movement into the Phase-Out ranges for both AFDC-FG and AFDC-U households. While a majority of AFDC-U households in the Replenishment Sample that were eligible for a credit had an AGI in the Phase-In region of the program (50 percent in Table 4c), AFDC-FG households were more evenly distributed across the Phase-In, Flat and Phase-Out regions of the program in 1994.

For those households that do claim the EITC, and are eligible for it, we find that the credit they actually receive varies from $908 to $1,534 depending upon the tax year and sample considered. (The latter estimates are displayed in column (i) of Tables 4a through 4c.) Compared to what they received from the AFDC program, the average credit households in the Original Sample received was 20 percent of their annual AFDC benefits among AFDC-FG households and 14 percent for AFDC-U households in 1993. For households in this sample, this ratio increased to 28 percent for those on AFDC-FG and 21 percent on AFDC-U in 1994. Among households in the Replenishment Sample, this same ratio was 44 percent for AFDC-FG households and 32 percent for those on AFDC-U in 1994. The higher ratios of credits received to AFDC benefits in 1994, compared to those in 1993, presumably reflect the increased generosity of the EITC program (see Table 1) and the improvement in the California economy (see Figure 2). We find that the average credit in 1993 was higher for AFDC-FG households compared to those on the AFDC-U program, although this ordering is reversed in 1994.

[31] The statistics reported in Tables 4a through 4c are based on data for the entire Analysis Sample. While not presented, the corresponding estimates for the households who were on AFDC at least part of the tax years do not differ greatly from those in the above tables. Also note that Tables 4a through 4c are based on using Criterion A to determine whether households were eligible for an EITC.

Based on the classification scheme presented in Figure 1, we also examined the AFDC payments received, EDD earnings and information from tax returns for 5 additional groups of households:

(ii) EITC-Eligibles, who did not claim the EITC;

(iii) EITC-Eligibles who did not file a tax return (or, more accurately we did not find a match for them with the IRS data on returns in California and surrounding states);

(iv) EITC-Ineligibles, who did claim the EITC;

(v) EITC-Ineligibles who did not claim the credit; and

(vi) EITC-Ineligibles who did not file a tax return.

Estimates for these groups are also found in Tables 4a through 4c. (The above numbering of groups corresponds to that for the columns in these tables.) While we postpone a discussion of the non-compliance aspects of the outcomes for Group (iv) until the following subsection, it is of interest to compare the differences in earnings, welfare benefits and tax return information for the remaining groups. Several patterns emerge. (We caution that the sample sizes for groups (ii) and (v) are typically very small.) With respect to the amount of AFDC received in a tax year, we find that groups which did not file tax returns, regardless of their EITC-eligibility status, typically received more in AFDC benefits than households in the other four groups. This pattern may reflect that households in these groups were either less able to generate earnings or, especially for those in Group (iii), that they perceived they might jeopardize the receipt of their AFDC benefits if they filed tax returns so as to claim the EITC, and, possibly, have to pay taxes on their earnings. There does not appear to be substantial differences in the AFDC benefits received across the remaining groups (compare columns (i), (ii), (iv) and (v)). The only exception to this pattern is for ineligible non-participants in our Replenishment Sample (column (v) in Table 4c) who received less in AFDC during the 1994 tax year and spent of that year on AFDC than did the other groups.

Among those groups that filed tax returns—Groups (i), (ii), (iv) and (v)—households in non-participant groups typically had substantially higher earnings than did households in the EITC participant groups. While the shares of our sample in either of the non-participant groups are typically small, it appears that households in these two groups, especially those in Group (v), often earned enough income so as not to qualify for a credit. For example, among Group (iv) households in the Original Sample in 1993 (column (v) in Table 4a) only 32 and 38 percent of AFDC-FG and AFDC-U households, respectively, had incomes below the EITC maximum income threshold.

Finally, we find that for households who were classified as being eligible for the EITC—Groups (i) and (ii)—households that claimed the EITC were more likely to use a paid tax preparer than were households who did not claim a credit. One would expect that paid preparers have a greater knowledge of the IRS rules, including those for the EITC. Furthermore, such returns are more likely to be filed electronically with the IRS than are ones filed in paper form and electronically filed returns have a shorter waiting time to receive a refund, including a tax credit. We note that among households that we classified as ineligible for a credit—Groups (iv) and

(v)—there is no clear pattern in the propensity for EITC claimants and non-claimants to use a paid tax preparer across our various samples, types of AFDC assistance, and tax years.[32]

3.4 Estimates of Rates of EITC Noncompliance

In the last three columns of Tables 3a and 3b, we present estimates of EITC non-compliance rates for our alternative samples, AFDC programs and tax years. Recall from the Introduction that recent evidence gathered by the IRS and GAO have raised serious concerns about the incidence of noncompliance associated with the EITC.[33] The most recent study of returns filed in 1995 for the 1994 tax year found that 25.8 percent of EITC claims exceeded the amount for which taxpayers were eligible and the corresponding rate was almost 33 percent based on an IRS study conducted for 1988 tax returns. Based on our data for households who are on, or have recently been on, AFDC in California, we find rates that are quite similar to those found in these previous studies for the nation as a whole. We estimate that between 17 and 57 percent of the AFDC-FG households in our sample that claimed an EITC were not eligible for it and that the corresponding range for AFDC-U households was between 9 and 45 percent. Using Criterion A for determining a household's EITC eligibility status, our estimates of noncompliance rates are around 25 percent. Furthermore, we find little evidence these rates vary much by AFDC type (AFDC-FG vs. AFDC-U) or by how long the household was on welfare (Original vs. Replenishment samples).

While our intermediate estimates of EITC noncompliance suggest that the incidence of this problem among a low-income, welfare population that claimed the EITC is not substantially higher than has been found for all EITC claimants in the nation, regardless of their participation in AFDC, it is important to emphasize the potential for bias in our estimates. Such bias would arise due to misclassification of households with respect to EITC eligibility status. With respect to the income component of EITC eligibility, we have noted that all of our eligibility criteria are biased toward understating income. This is because households are likely to understate their income to county welfare departments; furthermore, our other measure, based on earnings subject to California's UI and DI programs, undercounts self-employment income and income earned in jobs not covered by the UI insurance system. These shortcomings of our income measures would suggest that we are likely to *understate* a household's earnings from either EDD or reported Gross Earned Income (GEI) used for either AFDC or food stamp benefit determination. As such, it would appear that we are more likely to classify too many households as ineligible for EITC, at least based on meeting the income criteria of EITC eligibility.

There is a bias to the extent we classify households with self-employment income (and no wage income) as ineligible for the EITC. Based on our analysis of those who file tax returns in our samples, we find that very few claim this source of income (Schedule C income) on their returns. Not surprisingly, the highest percentages are seen in the category of ineligible EITC participants, i.e., Group (iv). As recorded in Tables 4a through 4c, between 9 and 15 percent of ineligible participant AFDC-FG households claimed self-employment income on their tax returns

[32] Preparer usage is highest among AFDC-U households in the Replenishment Sample. In particular, 79 percent of ineligible participants in this group used preparers.

[33] Both the chair of the U.S. House Ways and Means and Senate Finance Committees have recently held hearings on this issue.

while among ineligible participant AFDC-U households between 29 and 37 percent claimed it. The incidence of self-employment income in the other groups who filed returns—Groups (i), (ii), and (v)—is very low, never exceeding 10 percent of filers in a particular group. However, because few households are classified as being in the Ineligible, non-participant group (see the "% of Sample in Category" entry for column (v) in Tables 4a through 4c), the biases in our estimate EITC eligibility and noncompliance rates due to failing to count self-employment income correctly would be minimal. For example, if we reclassified all of ineligible, non-participant households with self-employment income to be eligible for the EITC—and used the same measures for qualifying children as in Criterion A—we would increase each of the "intermediate" estimates of EITC eligibility by 1 to 2 percentage points and reduce the corresponding noncompliance rates by between 2 to 3 percentage points. Thus, our eligibility and non-compliance rates do not appear to be terribly vulnerable to the failure to measure self-employment income, at least for our sample of current, or recent, AFDC recipients.

With respect to the presence-of-a-qualifying-child component, we have noted that our data are less than ideal, given that we must rely on evidence from county welfare records to identify children. Such counts of children present for purposes of determining AFDC benefits may, themselves, be subject to overreporting as households seek to increase the amount of the AFDC benefits. Furthermore, we have no information about whether the head of a household has a "qualifying" child in the household during months when the household is not on welfare. Finally, we have no other way of determining whether children are present other than using information from the welfare department records.

4. Comparability of Family Structure and Income Across Data Sources

In an effort to determine their reliability, we investigate the consistency of our various data sources on income and presence of children in households in this section. We begin by presenting comparisons of measures of the presence and number of children in AFDC households with the number of exemptions claimed in tax returns. As the discussion at the end of the previous section indicates, one of the problems with our analysis of EITC eligibility and noncompliance appears to be our difficulty in accurately measuring the presence of a qualifying child with the data we have from various administrative records systems. Clearly, this limitation chastens us from making strong claims about the robustness of our findings, especially with respect to rates of noncompliance. Furthermore, it is clear, especially with respect to measuring the qualifying child component of EITC eligibility, that we cannot completely resolve this issue with the data we have at hand. Nonetheless, we can examine the *consistency* of the measures across the data sources available to us in order to gain some sense of the *internal validity* of our findings. Accordingly, in this final section of our report, we present several comparisons of the information about children, and family structure in general, across our various data sources.

We also present comparisons of the income measures for the households in our sample from these various sources. Such a comparison provides us with a sense of the internal consistency of the measurement of income for disadvantaged households. Furthermore, such comparisons are of particular interest for learning more about the population from which our sample is drawn. This is especially true as federal and state governments attempt to redesign social programs to encourage work and minimize dependence on cash assistance. Such comparisons can help to improve our measurement of income and minimize potential misreporting and fraud. Ex-

amining income reporting for this population is of particular interest in the context of the AFDC and EITC programs because of the potentially conflicting incentives associated with these two programs. The AFDC program would appear to have encouraged households to underreport their wage and salary earnings while the EITC program gives households the opposite incentive in the Phase-In range of the credit. As such, measuring how the same households "report" to either administrative authority is of particular interest.

4.1 Household Size Comparisons

We employ two approaches for comparing family size reported to transfer and tax authorities. First, we compare the number of children age 18 and younger in the households in our sample. Recall that information on family composition is derived from AFDC case records and is subject to the problem that we do not know family composition during periods when these units are not receiving benefits. Nonetheless, we compare the resulting data on family composition from this source with the number of dependents claimed by the household on its tax return. More precisely, we estimate household size using monthly data on the number of AFDC-eligible children in a household from CWAD records. We identify the maximum number of eligible children in any one month of the year and compare this with the number of exemptions the household claims on its tax return(s). This comparison gives one measure of the degree to which household size differs across and tax and transfer records.[34]

Table 5a shows cross-tabulations of the number of children in an AFDC-FG household by the number of exemptions claimed on the household's tax return(s). In this table, we present these comparisons both for the Original Sample and for the Replenishment Sample. Table 5b shows similar information for AFDC-U households. As shown in the second to last columns of Tables 5a and 5b, the numbers of children reported to transfer and tax authorities are consistent across 57 to 65 percent of AFDC-FG households, and 74 to 80 percent of AFDC-U households. We observe the high estimate for AFDC-FG households in tax year 1993 and the low estimate in 1994, both for the Original Sample. For AFDC-U cases, however, around 75 percent of the reports between the two data sources agree in both 1993 and 1994 for the Original Sample. Lower match rates between AFDC records and exemptions for AFDC-FG households compared to AFDC-U households are not particularly surprising: the absent parent or other guardian may rightfully claim the child as a dependent for tax purposes. We do not see a consistent pattern of over- or under-reporting of family responsibilities to either tax or transfer authorities.

Another way of assessing the accuracy of the measures of the presence of children available in the CWAD data is to examine how a household's status as an AFDC-FG or AFDC-U case compares with the type of tax return(s) filed by adults in the household. Throughout a year, a household may maintain its designation as an AFDC-FG or AFDC-U case, switch from AFDC-FG to AFDC-U or vice versa, or switch to another type of assistance. We compare these classifications with the filing status reported on the household's tax return. Most AFDC-FG households

[34] We first adjust the number of exemptions to account for adults in the household. For example, we subtract one for a return with Head of Household filing status, and subtract two for a return with Joint filing status. The comparison of children in the AFDC case unit with the adjusted exemption count is admittedly a very rough comparison of household membership. Dependents for tax purposes do not necessarily reside with the filer. Furthermore, our data do not contain information on relationships among household members.

that file tax returns are likely to use the Head of Household filing status, whereas most AFDC-U households are likely to use the Married filing status. Noticeable departures from these expected patterns may indicate purposeful misreporting of household composition to tax or transfer authorities.

Table 6a presents comparisons of AFDC case status to type of tax return filed by the household, for households that entered the sample as AFDC-FG cases. Table 6b displays similar information for households that entered the sample as AFDC-U cases. First, we comment on tax return filing rates for these types of households. Filing rates are highest for households that move from AFDC to another type of assistance by the end of the year. Additionally, households that are AFDC-U households at the end of the year, no matter whether they were AFDC-U or AFDC-FG when drawn into the sample, generally have higher filing rates than those households that switched to or continued as AFDC-FG cases. Finally, Replenishment households that were AFDC-U households when drawn into the sample have higher overall filing rates than those in the Original Sample for all types of cases.

Of AFDC-FG households that remain AFDC-FG throughout the year and that file returns, most—between 62 and 68 percent—file as Heads of Household. The next most frequent filing status observed among this group is the Married status (*not* Married Filing Separately), then Single. These AFDC-FG households that use the Married filing status may or may not be reporting different family structure to tax and transfer authorities: we know that some women in AFDC-FG households are married or separated and thus filing a tax return using the Married filing status does not necessarily indicate fraudulent behavior.

The majority of households that entered the sample as AFDC-FG and moved to AFDC-U by the end of the year use the Married filing status, though many file as Heads of Household. Not surprisingly, almost all households that entered the sample and remain as AFDC-U and that file tax returns use the Married filing status. Most AFDC-U households that move to AFDC-FG or other case statuses file as Married or Head of Household, though we do see some filing as Single. Overall, these rough comparisons do not highlight large differences in reporting of family structure to tax and welfare authorities.

4.2 Match Rates and Income Comparisons

The linked administrative records from AFDC files and the Base Wage File, together with individual tax returns, provide a unique opportunity to examine match rates and consistency of income reports, and to determine whether reports of income to the various sources are consistent with incentives. We know that incentives exist to underreport income to transfer authorities to maximize benefits, to overreport income to tax authorities in the phase-in range of the EITC, and to underreport income to tax authorities in the phase-out range of the credit. There are fewer incentives for employers to misreport earnings to the UI/DI system, as discussed in Section 2 of this report, so employer-reported income in the Base Wage File can be used as the basis of comparisons to AFDC income reports and tax records.

4.2.1 Wage Earnings from EDD Base Wage File versus AFDC Gross Earned Income in the CWAD File [35]

We begin by describing the matching process used to find the adults from our CWPDP Analysis Sample in the EDD Base Wage File. The latter file contains quarterly wage information for individuals who worked in jobs in California that are covered by the UI and DI system over the period from 1984 through 1994. We first ascertained whether the adults in the CWPDP sample had a record in the EDD Base Wage File for any quarter in this span of years, where this match was based on Social Security numbers. Note that an "EDD Match" would occur even if an adult in our sample worked in the UI/DI covered sector at some point during this period but did not work at all in 1993 or 1994, or did not work in a covered job in those two years. In either case, the adult would have an EDD Match but their EDD wage earnings would be recorded as zero during the particular year. Conditional on a match with the EDD data, we obtain a measure of a household's annual earnings by aggregating Base Wage File quarterly income reports for all adults in the household.

We note that earnings information is available in the Base Wage File regardless of whether the adult or household is eligible for AFDC in any particular month, quarter, or year. No record of the household's earnings is available in the CWPDP analysis file, however, during months the household is ineligible for AFDC. Thus, simply comparing a household's annual earnings in the Base Wage File with the sum of monthly earnings in its AFDC case file, regardless of the number of months of AFDC eligibility, is not a meaningful comparison.

To accurately assess the variations between income in the Base Wage File and in AFDC records, we must impose a sample restriction and use only those households that are eligible for AFDC during all twelve months of the year. We group income into $2,500 dollar increments. Even though this is quite a wide band of income, we expect to see some discrepancies in income reports across the different data sources due to the slightly different measures of income that each dataset uses, and to the coverage and reporting issues, discussed in previous sections of this report.

Match rates for adults from the CWPDP Analysis Sample in the EDD Base Wage File were quite high, for both AFDC-FG and AFDC-U households: We obtained matches to Base Wage File records for 78 percent of adults in AFDC-FG households and 75 percent of adults in AFDC-U households. While these rates are high, it is important to keep in mind that a match does not indicate earnings in a particular year, and that the sample submitted for a match is a restricted sample (see the Appendix).

Table 7a presents comparisons of income reports from the EDD Base Wage File and reported AFDC Gross Earned Income (GEI) for AFDC-FG households obtained from the CWAD. Yearly comparisons are presented for the Original Sample and for the Replenishment Sample. Table 7b shows similar information for AFDC-U households. Between 66 to 75 percent of AFDC-FG households and 49 to 65 percent of AFDC-U households in the restricted sample

[35] We also compared Base Wage File earnings with earnings reported for food stamp eligibility determination. The reporting patterns of food stamp income were quite similar to reports of income for AFDC eligibility determination. We present findings only on the comparison of Base Wage File and AFDC gross earned income.

show zero earnings in the Base Wage File and zero income in AFDC administrative records.

The cases of interest for our purposes, however, are the ones in which an adult in a household has positive earnings in the Base Wage File but has *underreported* GEI to AFDC. Conditional on having positive income in the Base Wage File in a given year, we find that 55 to 69 percent of AFDC-FG households and 36 to 41 percent of AFDC-U households appear to underreport income to AFDC authorities. These estimates reflect a conservative estimate of underreporting to AFDC authorities for three reasons. First, some households have positive earnings in the Base Wage File but have a missing value recorded for AFDC Gross Earned Income in the AFDC case files. We attribute this to administrative error (since the records show the household eligible for AFDC), not to purposeful underreporting by the household. Second, the AFDC income measure may include income for other persons in the household for whom we do not have access to Base Wage File earnings records. Finally, as discussed in Section 2, AFDC GEI is a more inclusive measure of income than earnings in the Base Wage File. Nonetheless, the evidence summarized in Tables 7a and 7b seems to clearly indicate that households on AFDC underreport their income to county welfare departments, even earnings that are readily verifiable.

4.2.2 Wage Earnings from EDD Base Wage File versus Income data from Tax Returns for Our Analysis Sample

The California Franchise Tax Board matched records from the Base Wage File with federal tax returns from 1993 and 1994. Most people who have no earnings in the Base Wage File are not likely to file tax returns. Some with zero earnings will file returns, however, if they are self-employed or work in sectors that are not covered by UI/DI. Even still, families that have earnings but do not have positive tax liabilities are not required to file tax returns. A married couple, for example, is not required to file if their income is below the standard deduction and two exemptions ($12,200 in 1997), regardless of how many children they have. Other families may be required to file, but choose not to do so.

Given matches, we also compare aggregate quarterly earnings in the Base Wage File to two items on tax returns: wage and salary earnings reported on Line 7 of an individual's 1040 form and their Adjusted Gross Income (AGI, reported on line 31 of the 1040 form). We would expect earnings in the Base Wage File and Line 7 Wage and Salary earnings to correspond quite closely because they are essentially the same measure of income. Similarly, Base Wage File earnings and AGI should not differ greatly because the additions and adjustments to income that affect AGI are not likely to be heavily used by persons in our sample.[36]

Examining the match rates for filing tax returns with those assistance units in our CWAD subsample, we find that between 22 to 33 percent of AFDC-FG households and 31 to 49 percent of AFDC-U households in our sample filed tax returns (see Tables 3a and 3b). Match rates between adults in the CWPDP Analysis File and federal tax returns were higher for AFDC-U households than AFDC-FG households, higher in 1994 than in 1993 for households in the Original Sample, and higher for both AFDC-U and AFDC-FG households in the Replenishment Sample compared to AFDC-U and AFDC-FG households in the Original Sample. We find that 22

[36] Examples of these additions and adjustments to income include dividend income, alimony, capital gains, and IRA and Keogh deductions.

percent of AFDC-FG households and 31 percent of AFDC-U households in the Original Sample filed returns in 1993. The filing rates for these samples increased to 31 and 37 percent, respectively, in tax year 1994. In the Replenishment Sample, 33 percent of AFDC-FG households and 49 percent of AFDC-U households filed returns in 1994.

Tables 8a and 8b display summary statistics for income comparisons for AFDC-FG and AFDC-U households, respectively, of income in the EDD Base Wage File and income reported as AGI on the federal tax return. Tables 9a and 9b show similar information for comparisons of EDD Base Wage File earnings and Line 7 Wage and Salary income on the federal tax return. Comparisons of earnings in the Base Wage File to tax return data show, as we expect, that most households with zero earnings in the Base Wage File do not file tax returns. Conditional on having positive earnings in the Base Wage File, we find very low rates of underreporting of income to tax authorities. Only 3 to 6 percent of AFDC-FG and AFDC-U households report lower AGI than we find in the Base Wage File. Similarly, only 4 to 5 percent of AFDC-FG and AFDC-U households in the restricted sample report lower Line 7 earnings than we detect in Base Wage File records. Tables 8a through 9b highlight another one important aspect of EITC participation: most households that have positive earnings but do not file returns (and thus do not claim the EITC) have *very low* earnings ($2,500 or less). We observe these households on the "no tax return" rows under AGI and Line 7 comparisons.

5. Conclusion

There appear to be several conclusions one can draw, albeit tentatively, from the findings of this study about the role of the EITC and with the incentives that programs like AFDC and the EITC create for reporting and behavior. First, we find that a non-trivial proportion of households on AFDC were eligible for and claimed the earned income tax credit during the first half of the 1990s. In particular, we find that between 12 to 20 percent of the households in our sample on AFDC-FG and between 20 and 33 percent of those on the AFDC-U program actually received a earned income tax credit for years in which they were also receiving AFDC. While not huge, this degree of overlap between the utilization of these two programs is not negligible. Furthermore, while the average annual credit received by households in our sample is small—between $166 to $381 per year for those on AFDC-FG and between $236 to $615 for those on AFDC-U—this amount increased from 1993 to 1994, amounting to between 5 to 12 percent of the average annual benefits these households received from the AFDC program. In short, this study documents an important, and non-negligible, overlap between the AFDC and EITC programs.

Our evidence on the EITC "involvement" of those on welfare also provides some encouragement that the provisions of the recently passed Personal Responsibility and Work Opportunities Act which seek to encourage greater work among this population may succeed. These findings are somewhat surprising for at least two reasons. First, California experienced a severe recession during this period; one would have expected that the employment opportunities of low-income adults would have been severely limited. Second, most previous studies have found that the rates of labor force participation among those on AFDC are very low. The basis for the latter conclusion is the incidence of labor market earnings reported to welfare departments by AFDC recipients. Our study documents, as have others, that these reports severely underreport such income. Moreover, we find that this underreporting is not simply attributable to the failure to report "off-the-books" earnings from casual labor, as suggested in a recent ethnographic study of

welfare recipients by Kathryn Edin. In fact, we find that earnings from employment that is verifiable and readily documented are substantially underreported as well.

References

Bane, M. J. and D. Ellwood (1983), *The Dynamics of Dependence: The Routes to Self-Sufficiency*, Report to DHHS/ASPE, Cambridge, MA: Urban Systems Research and Engineering, Inc., June 1983.

Becerra, R., A. Lewin, M. Mitchell, and H. Ono (1996), *California Work Pays Demonstration Project: Interim Report of First Thirty Months*, School of Public Policy and Social Research, University of California, Los Angeles, December 1996.

Blank, Rebecca M. and Patricia Ruggles (1993), "When Do Women Use AFDC and Food Stamps: The Dynamics of Eligibility versus Participation," Northwestern University and the Urban Institute. Washington, D.C. June 1993.

Duncan, A. and C. Giles (1996), "Labour Supply Incentives and Recent Family Credit Reforms," *Economic Journal*, January, 142-155

Edin, K. (1995), "The Myths of Dependence and Self-Sufficiency: Women, Welfare, and Low-Wage Work," *Focus*, Madison, WI: Institute for Research on Poverty, 17(2), pp. 1-9.

Ellwood, D. (1986), "Targeting the Would-Be Long Term Recipient of AFDC: Who Should be Served?" Princeton, NJ: Mathematica Policy Research, January 1986.

Liebman, J. (1995a), Who are the Ineligible EITC Recipients?" Harvard University and NBER, Unpublished manuscript, October 1995.

Liebman, J. (1995b), "Noncompliance and the EITC: Taxpayer Error or Taxpayer Fraud?" November 22, mimeo, Harvard University and NBER, Unpublished manuscript, October 1995.

Gritz, R. M., and T. MaCurdy (1992), "Patterns of Welfare Utilization and Multiple Program Participation among Young Women," Stanford, CA: The Hoover Institution, January 1992.

O'Neill, J., D. Wolf, L. Bassi, and M. Hannan (1984), *An Analysis of Time on Welfare*, Report submitted to DHHS/ASPE, Washington, DC: The Urban Institute, June 1984.

Scholz, J. K. (1994), "The Earned Income Tax Credit: Participation, Compliance, and Anti-poverty Effectiveness," *National Tax Journal*, March, 59-81.

Spalter-Roth, R., B. Burr, H. Hartmann, and L. Shaw (1994), *Welfare that Works: The Working Lives of AFDC Recipients*, Washington, DC: Institute for Women's Policy Research.

U. S. General Accounting Office (1996a), *Earned Income Credit: Profile of Tax Year 1994 Credit Recipients*, GAO Briefing Report, June 1996.

U. S. General Accounting Office (1996b), *Tax Administration: Tax Compliance of Nonwage Earners*, GAO Report, August 1996.

U.S. House of Representatives, Committee on Ways and Means (1996), *Green Book: Background Material and Data on Programs within the Jurisdiction of the Committee on Ways and Means,* Washington, DC: Government Printing Office.

Figure 1

Earned Income Tax Credit: 1995

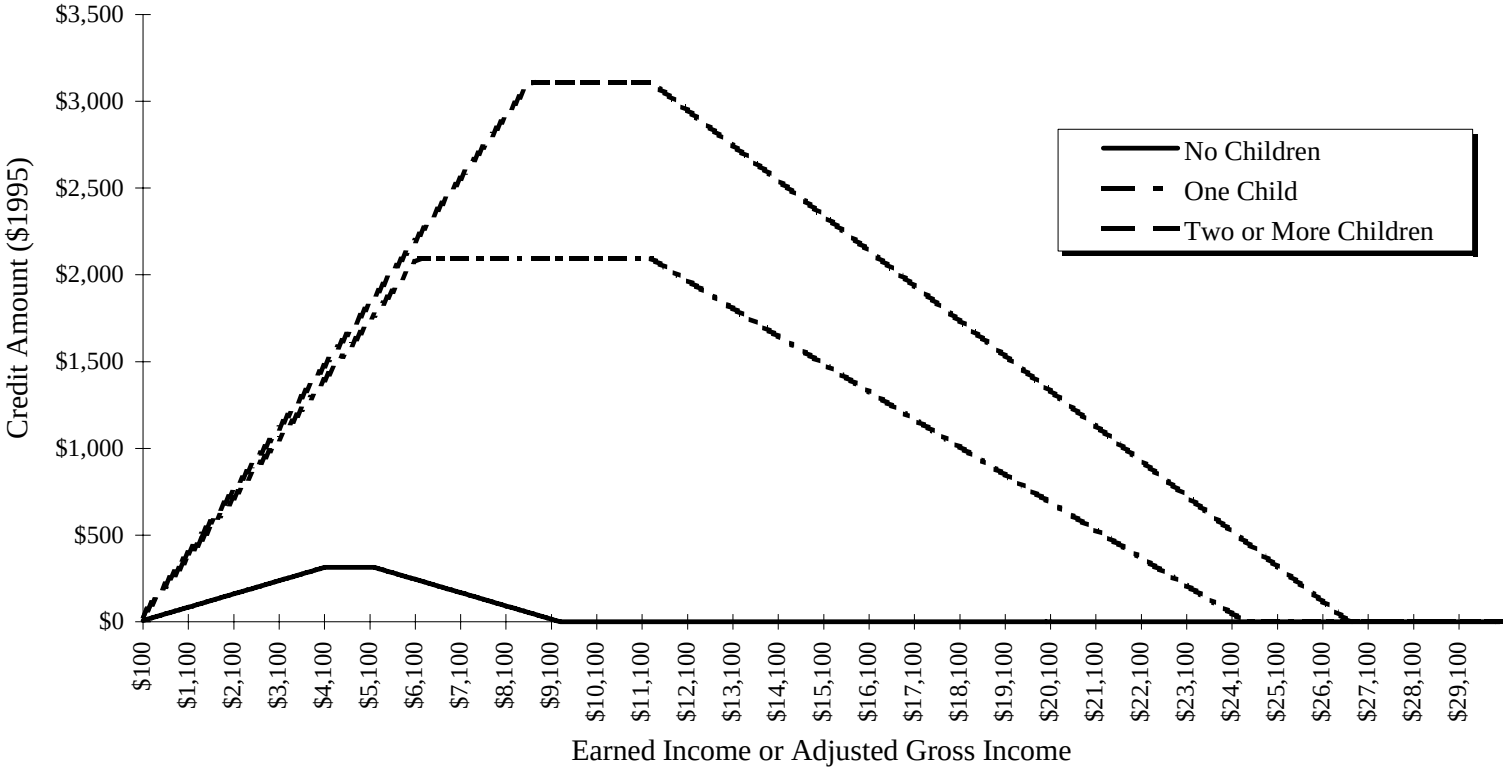


Figure 2

Annual Growth Rates in Employment: All Industries, 1986-1994

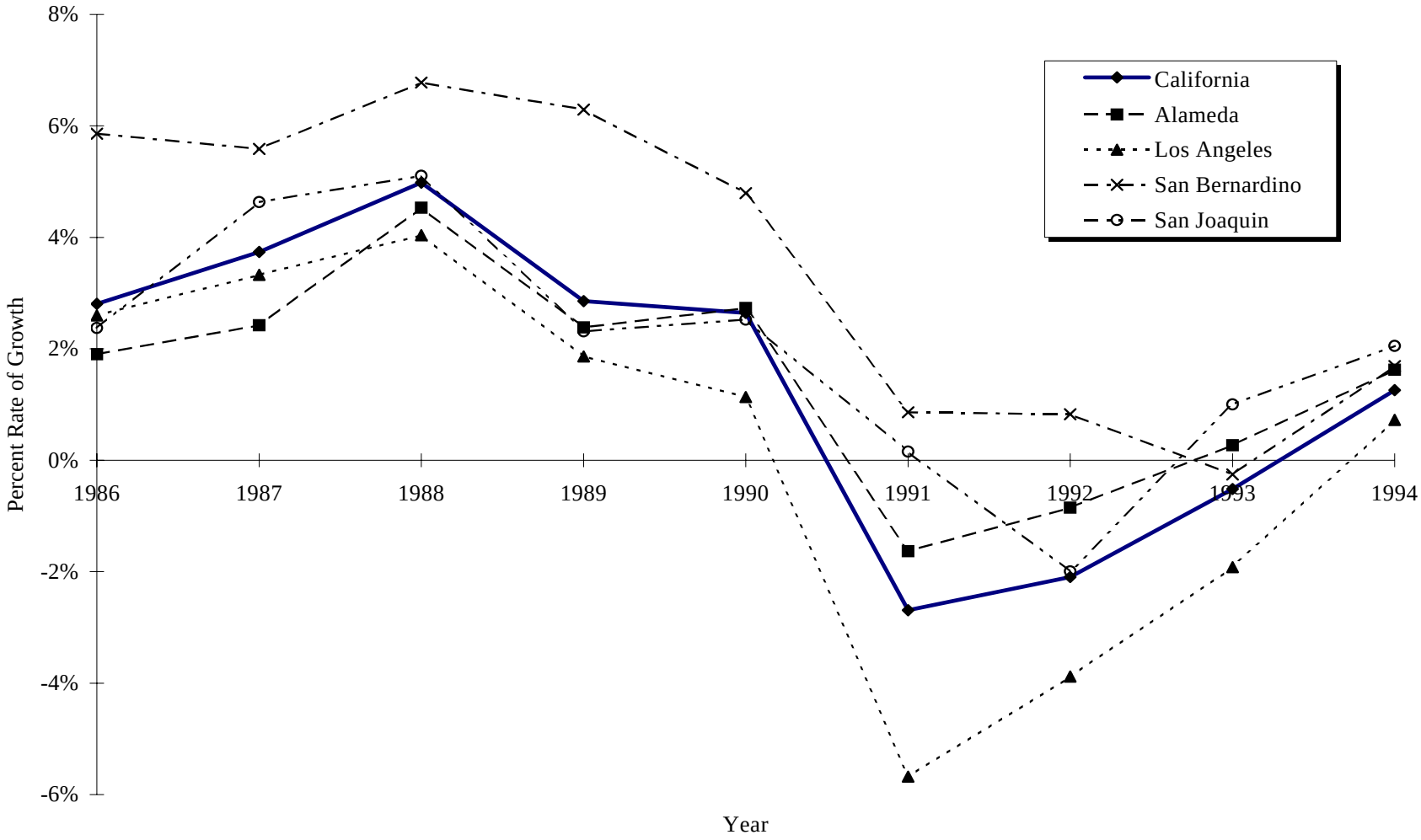


Figure 3
Schematic Representation of Classification of Assistance Units with Respect to EITC Eligibility and Participation

Tax Return and EITC Filing Status	*Meets Candidate Criteria Used for EITC Eligibility*	*Does Not Meet Candidate Criteria Used for EITC Eligibility*
Tax Return found for Head(s) of AU and EITC claimed	**Cell I:** Eligible participant	**Cell IV:** Ineligible participant
Tax Return found for Head(s) of AU and EITC not claimed	**Cell II:** Eligible non-participant	**Cell V:** Ineligible non-participant
No Tax Return found for Head(s) of AU	**Cell III:** Eligible non-participant	**Cell VI:** Ineligible non-participant

Table 1
EITC Program Parameters, Selected Years (dollar amounts unadjusted for inflation)

Year	Credit Rate	Flat Region		Maximum Credit	Phase-out Region	
		Beginning	Ending		Phase-out Credit	Income
1993						
1 child	18.5%	$7,750	$12,200	$1,434	13.21%	$23,050
2+ kids	19.5%	$7,750	$12,200	$1,511	13.93%	$23,050
Young child[1]	5.0%	$7,750	$12,200	$388	3.57%	$23,050
Health credit[2]	6.0%	$7,750	$12,200	$465	4.285%	$23,050
1994						
1 child	26.3%	$7,750	$11,000	$2,038	15.98%	$23,755
2+ Kids	30.0%	$8,425	$11,000	$2,528	17.68%	$25,296
No child	7.65%	$4,000	$5,000	$306	7.65%	$9,000
1997						
1 child	34.0%	$6,500	$11,930	$2,210	15.98%	$25,760
2+ kids	40.0%	$9,140	$11,930	$3,656	21.06%	$29,290
No child	7.65%	$4,340	$5,430	$332	7.65%	$9,770

Source: Green Book 1996, and Office of Tax Policy, U.S. Treasury Department.

Table 2a
Background Characteristics for CWPDP Analysis Sample
Original Sample, Measured in Tax Year 1993

Variables	County				
	Alameda	*Los Angeles*	*San Bernardino*	*San Joaquin*	*Four Counties*
AFDC-FG Cases					
Race:					
White	0.18	0.19	0.47	0.41	0.24
Black	0.61	0.41	0.22	0.12	0.39
Hispanic	0.10	0.33	0.28	0.23	0.29
Asian	0.11	0.07	0.03	0.23	0.07
Gender: (fraction male)	0.07	0.09	0.09	0.14	0.09
Numbers of Adults in AU:					
1	0.98	0.94	0.93	0.95	0.94
2	0.02	0.06	0.07	0.05	0.06
Numbers of Children in AU:					
0	0.03	0.01	0.02	0.02	0.02
1	0.41	0.39	0.33	0.33	0.38
2	0.30	0.30	0.33	0.32	0.30
3+	0.25	0.30	0.31	0.34	0.30
No. of Adults with EDD Earnings:					
0	0.69	0.75	0.73	0.71	0.74
1	0.30	0.25	0.26	0.29	0.26
2	0.01	0.00	0.01	0.01	0.00
No. of Assistance Units	1,584	2,329	1,567	556	6,036
AFDC-U Cases					
Race:					
White	0.28	0.47	0.52	0.31	0.46
Black	0.12	0.05	0.09	0.04	0.06
Hispanic	0.07	0.23	0.28	0.12	0.22
Asian	0.51	0.25	0.09	0.51	0.25
Gender: (Fraction Male)	0.49	0.49	0.48	0.49	0.49
Numbers of Adults in AU:					
1	0.08	0.19	0.16	0.07	0.17
2	0.92	0.81	0.84	0.93	0.83
Numbers of Children in AU:					
0	0.03	0.00	0.02	0.00	0.01
1	0.15	0.18	0.16	0.13	0.17
2	0.26	0.41	0.29	0.22	0.37
3+	0.55	0.40	0.53	0.65	0.45
No. of Adults with EDD Earnings:					
0	0.76	0.71	0.62	0.62	0.69
1	0.20	0.25	0.30	0.31	0.26
2	0.04	0.04	0.08	0.06	0.05
No. of Assistance Units	750	1,308	766	221	3,045

Table 2b
Background Characteristics for CWPDP Analysis Sample
Replenishment Sample, Measured in Tax Year 1994

Variables	County				
	Alameda	*Los Angeles*	*San Bernardino*	*San Joaquin*	*Four Counties*
AFDC-FG Cases					
Race:					
White	0.28	0.30	0.48	0.54	0.35
Black	0.45	0.20	0.18	0.14	0.21
Hispanic	0.15	0.44	0.31	0.20	0.38
Asian	0.12	0.04	0.01	0.10	0.04
Gender: (fraction male)	0.07	0.07	0.15	0.13	0.09
Numbers of Adults in AU:					
1	0.94	0.97	0.94	0.95	0.96
2	0.06	0.03	0.06	0.05	0.04
Numbers of Children in AU:					
0	0.27	0.24	0.27	0.37	0.25
1	0.35	0.47	0.36	0.37	0.43
2	0.25	0.18	0.22	0.15	0.20
3+	0.12	0.11	0.15	0.10	0.12
No. of Adults with EDD Earnings:					
0	0.50	0.66	0.60	0.49	0.63
1	0.47	0.34	0.40	0.47	0.36
2	0.02	0.00	0.00	0.03	0.01
No. of Assistance Units	139	232	253	46	670
AFDC-U Cases					
Race:					
White	0.29	0.28	0.49	0.30	0.34
Black	0.17	0.06	0.07	.	0.07
Hispanic	0.25	0.52	0.37	0.53	0.47
Asian	0.29	0.14	0.06	0.15	0.13
Gender: (Fraction Male)	0.49	0.51	0.50	0.48	0.51
Numbers of Adults in AU:					
1	0.30	0.54	0.40	0.26	0.49
2	0.70	0.46	0.60	0.74	0.51
Numbers of Children in AU:					
0	0.24	0.20	0.26	0.58	0.23
1	0.25	0.29	0.20	0.02	0.26
2	0.44	0.33	0.25	0.15	0.31
3+	0.08	0.18	0.29	0.24	0.21
No. of Adults with EDD Earnings:					
0	0.55	0.59	0.38	0.20	0.52
1	0.36	0.35	0.47	0.53	0.39
2	0.09	0.06	0.15	0.28	0.09
No. of Assistance Units	83	215	219	35	552

Table 3a
EITC Eligibility, Participation and Non-Compliance Rates: Lower and Upper Bound and "Intermediate" Estimates
AFDC-FG and AFDC-U Households, Tax Years 1993 and 1994
(All Households in Sample, including those who went off AFDC)

Sample	EITC Eligibility Rate			% with Tax Return	EITC Participation Rate[1]			EITC Non-Compliance Rate[2]		
	Lower Bound	*Intermediate Estimate*	*Upper Bound*		*Lower Bound*	*Intermediate Estimate*	*Upper Bound*	*Lower Bound*	*Intermediate Estimate*	*Upper Bound*
AFDC-FG Households:										
Original Sample, Tax Year 1993	21.97[e]	26.02[a]	29.05[b]	22.46	42.21[f]	46.73[a]	46.73[a]	22.19[b]	26.48[a]	40.58[e]
Original Sample, Tax Year 1994	23.22[e]	33.23[a]	37.17[b]	30.52	44.55[f]	49.83[a]	49.83[a]	17.85[b]	23.40[a]	49.53[e]
Replenishment Sample, Tax Year 1994	21.56[e]	36.08[a]	40.94[b]	32.93	47.56[f]	54.19[a]	54.36[c]	23.05[b]	27.08[a]	57.49[e]
AFDC-U Households:										
Original Sample, Tax Year 1993	25.64[e]	29.56[a]	37.71[b]	30.59	61.39[f]	67.46[a]	67.59[c]	11.89[b]	25.46[a]	36.15[e]
Original Sample, Tax Year 1994	25.90[e]	35.03[a]	44.23[b]	36.57	59.66[f]	64.77[a]	65.29[e]	11.86[b]	25.09[a]	44.15[e]
Replenishment Sample, Tax Year 1994	27.56[e]	41.17[a]	52.99[b]	49.04	71.86[b]	76.85[a]	83.82[e]	9.59[b]	24.88[a]	45.14[e]

Notes:

1. Participation rate is the percentage of "eligible" assistance units who claimed the EITC, where eligibility definition is noted for each entry.
2. Non-compliance rate is the percentage of assistance units that claimed the EITC who did not meet the eligibility definition, where the definition is noted for each entry.

EITC Eligibility Criteria Definitions:

Criteria	*Income Source*	*Child Eligible for AFDC*
Criterion A	UI/DI Wage Earnings > 0	0+ months during year
Criterion B	Max{UI/DI, Food Stamp AGI, AFDC AGI} > 0	0+ months during year
Criterion C	UI/DI Wage Earnings > 0	1+ months during year
Criterion D	Max{UI/DI, Food Stamp AGI, AFDC AGI} > 0	1+ months during year
Criterion E	UI/DI Wage Earnings > 0	6+ months during year
Criterion F	Max{UI/DI, Food Stamp AGI, AFDC AGI} > 0	6+ months during year

Table 3b
EITC Eligibility, Participation and Non-Compliance Rates: Lower and Upper Bound and "Intermediate" Estimates
AFDC-FG and AFDC-U Households, Tax Years 1993 and 1994
(Households on AFDC for at least one month in Tax Year)

Sample	**EITC Eligibility Rate**			**% with Tax Return**	**EITC Participation Rate[1]**			**EITC Non-Compliance Rate[2]**		
	Lower Bound	*Intermediate Estimate*	*Upper Bound*		*Lower Bound*	*Intermediate Estimate*	*Upper Bound*	*Lower Bound*	*Intermediate Estimate*	*Upper Bound*
AFDC-FG Households:										
Original Sample, Tax Year 1993	22.46[e]	25.87[a]	28.96[b]	22.22	42.45[f]	46.85[a]	46.85[a]	22.52[d]	26.81[a]	39.17[e]
Original Sample, Tax Year 1994	28.29[e]	31.90[a]	36.78[b]	27.53	45.68[f]	48.68[a]	48.68[a]	17.06[b]	24.28[a]	33.84[e]
Replenishment Sample, Tax Year 1994	28.43[e]	37.73[a]	44.15[b]	34.39	47.67[f]	53.51[a]	54.36[c]	23.03[b]	28.10[a]	46.35[e]
AFDC-U Households:										
Original Sample, Tax Year 1993	28.13[e]	31.48[a]	39.54[b]	31.98	58.98[f]	64.04[a]	64.04[c,a]	11.12[b]	24.49[a]	33.67[e]
Original Sample, Tax Year 1994	34.57[e]	37.98[a]	47.91[b]	38.01	57.28[f]	61.53[a]	61.53[a]	8.88[b]	23.10[a]	30.44[e]
Replenishment Sample, Tax Year 1994	36.43[e]	42.61[a]	57.20[d,b]	52.39	71.91[d,b]	77.96[a]	80.70[e]	8.44[b]	26.05[a]	34.54[e]

Notes:

1. Participation rate is the percentage of "eligible" assistance units who claimed the EITC, where eligibility definition is noted for each entry.
2. Non-compliance rate is the percentage of assistance units that claimed the EITC who did not meet the eligibility definition, where the definition is noted for each entry.

EITC Eligibility Criteria Definitions:

Criteria	*Income Source*	*Child Eligible for AFDC*
Criterion A	UI/DI Wage Earnings > 0	0+ months during year
Criterion B	Max{UI/DI, Food Stamp AGI, AFDC AGI} > 0	0+ months during year
Criterion C	UI/DI Wage Earnings > 0	1+ months during year
Criterion D	Max{UI/DI, Food Stamp AGI, AFDC AGI} > 0	1+ months during year
Criterion E	UI/DI Wage Earnings > 0	6+ months during year
Criterion F	Max{UI/DI, Food Stamp AGI, AFDC AGI} > 0	6+ months during year

Table 4a
AFDC and Tax Return Outcomes, By Status of Tax Match, Eligibility, and EITC Participation
Original Sample; Tax Year 1993[a]

Characteristic	*EITC-Eligible:*			*EITC-Ineligible:*			*All Households*
	EITC Participants (i)	*EITC Non-Participants* (ii)	*Did Not Have IRS Match* (iii)	*EITC Participants* (iv)	*EITC Non-Participants* (v)	*Did Not Have IRS Match* (vi)	
AFDC-FG Cases:							
% of Sample in Category:	12.16	2.81	11.05	4.38	3.11	66.49	100.00
Months on AFDC During Year:							
0	0.01	0.02	0.01	0.01	0.02	0.01	0.01
1 to 4	0.18	0.17	0.07	0.08	0.23	0.06	0.08
5 to 8	0.11	0.09	0.10	0.14	0.07	0.06	0.08
9 to 12	0.70	0.72	0.83	0.78	0.69	0.87	0.83
Annual AFDC Benefits Received:	$4,888	$5,130	$6,198	$6,447	$5,444	$6,824	$6,412
Average EDD Wage Earnings:	$6,573	$4,484	$1,893	$556	$2,064	$29	$1,242
Tax Filing Status:							
Single	0.08	0.63		0.03	0.09		0.07[b]
Married but Filed Separately	0.00	0.02		0.01	0.08		0.00[b]
Married	0.10	0.23		0.37	0.58		0.17[b]
Head of Household	0.81	0.12		0.59	0.25		0.75[b]
Tax Preparer:							
Was Tax Preparer Used?	0.54	0.52		0.58	0.46		0.55[b]
Adjusted Gross Income:	$7,676	$11,320		$9,179	$23,306		$2,378
Line 7, Wages and Salary:	$7,325	$10,195		$8,276	$20,890		$2,189
AGI income in EITC Range:							
Phase-In	0.59	0.60		0.46	0.17		0.56[b]
Flat	0.24	0.13		0.23	0.04		0.24[b]
Phase-Out	0.18	0.06		0.29	0.11		0.21[b]
Self-Employment Income:							
Fraction who Reported Any	0.03	0.02		0.13	0.09		0.06[b]
Was Tax Preparer Used?	0.58	1.00		0.73	0.57		0.62[b]
Ave Amt	$5,225	$11,370		$4,345	$8,726		$4,992[b]
EITC Received:	$988			$1,058			$166
AFDC-U Cases:							
% of Sample in Category:	19.94	1.90	7.72	6.81	1.94	61.69	100.00
Months on AFDC During Year:							
0	0.01	0.00	0.00	0.00	0.03	0.00	0.00
1 to 4	0.11	0.10	0.08	0.13	0.31	0.04	0.07
5 to 8	0.11	0.16	0.10	0.04	0.17	0.03	0.06
9 to 12	0.77	0.74	0.82	0.83	0.49	0.92	0.86
Annual AFDC Benefits Received:	$6,683	$6,523	$7,606	$7,322	$4,971	$8,895	$8,126
Average EDD Wage Earnings:	$5,936	$6,302	$2,277	$1,852	$8,667	$30	$1,792
Tax Filing Status:							
Single	0.01	0.32		0.01	0.06		0.03[b]
Married but Filed Separately	0.00	0.20		0.00	0.10		0.02[b]
Married	0.87	0.44		0.89	0.77		0.84[b]
Head of Household	0.11	0.05		0.09	0.06		0.10[b]
Tax Preparer:							
Was Tax Preparer Used?	0.59	0.37		0.64	0.51		0.58[b]
Adjusted Gross Income:	$7,118	$11,032		$7,522	$23,385		$2,595
Line 7, Wages and Salary:	$6,489	$9,848		$5,694	$22,781		$2,311
AGI income in EITC Range:							
Phase-In	0.66	0.61		0.67	0.27		0.63[b]
Flat	0.20	0.20		0.16	0.00		0.18[b]
Phase-Out	0.14	0.07		0.17	0.11		0.14[b]
Self-Employment Income:							
Fraction who Reported Any	0.06	0.09		0.29	0.10		0.12[b]
Was Tax Preparer Used?	0.63	1.00		0.73	0.36		0.66[b]
Ave Amt	$5,043	$5,581		$4,945	$1,936		$4,858[b]
EITC Received:	$908			$814			$236

Notes: a. EITC Eligibility Criterion A used.
b. Averages for those households which filed a tax return.

Table 4b
AFDC and Tax Return Outcomes, By Status of Tax Match, Eligibility, and EITC Participation
Original Sample; Tax Year 1994[a]

Characteristic	*EITC-Eligible:* *EITC Participants* (i)	*EITC Non-Participants* (ii)	*Did Not Have IRS Match* (iii)	*EITC-Ineligible:* *EITC Participants* (iv)	*EITC Non-Participants* (v)	*Did Not Have IRS Match* (vi)	*All Households*
AFDC-FG Cases:							
% of Sample in Category:	16.56	4.45	12.22	5.06	4.45	57.25	100.00
Months on AFDC During Year:							
0	0.05	0.06	0.03	0.02	0.05	0.01	0.02
1 to 4	0.17	0.16	0.10	0.12	0.21	0.10	0.12
5 to 8	0.15	0.20	0.11	0.15	0.13	0.10	0.12
9 to 12	0.63	0.57	0.75	0.71	0.61	0.79	0.74
Annual AFDC Benefits Received:	$4,568	$4,564	$5,728	$5,630	$4,732	$6,321	$5,774
Average EDD Wage Earnings:	$7,248	$5,393	$2,387	$1,367	$2,797	$101	$1,983
Tax Filing Status:							
Single	0.13	0.49		0.07	0.08		0.17[b]
Married but Filed Separately	0.00	0.04		0.00	0.07		0.02[b]
Married	0.12	0.28		0.37	0.47		0.24[b]
Head of Household	0.75	0.18		0.56	0.38		0.58[b]
Tax Preparer:							
Was Tax Preparer Used?	0.56	0.48		0.57	0.62		0.56[b]
Adjusted Gross Income:	$9,312	$16,857		$10,772	$33,737		$4,339
Line 7, Wages and Salary:	$8,966	$16,137		$9,684	$32,695		$4,148
AGI income in EITC Range:							
Phase-In	0.53	0.45		0.42	0.05		0.50[b]
Flat	0.12	0.06		0.10	0.02		0.11[b]
Phase-Out	0.34	0.13		0.46	0.17		0.37[b]
Self-Employment Income:							
Fraction who Reported Any	0.03	0.03		0.15	0.07		0.06[b]
Was Tax Preparer Used?	0.78	1.00		0.69	0.94		0.82[b]
Ave Amt	$4,690	$4,651		$5,364	$9,715		$5,529[b]
EITC Received:	$1,294			$1,282			$279
AFDC-U Cases:							
% of Sample in Category:	22.69	2.59	9.75	7.60	3.69	53.68	100.00
Months on AFDC During Year:							
0	0.06	0.08	0.03	0.01	0.04	0.01	0.03
1 to 4	0.13	0.17	0.13	0.14	0.24	0.07	0.10
5 to 8	0.11	0.12	0.14	0.12	0.20	0.06	0.09
9 to 12	0.70	0.62	0.71	0.72	0.51	0.86	0.78
Annual AFDC Benefits Received:	$6,292	$5,350	$6,908	$6,897	$5,291	$8,426	$7,482
Average EDD Wage Earnings:	$6,620	$9,308	$3,446	$3,148	$11,709	$214	$2,865
Tax Filing Status:							
Single	0.04	0.27		0.02	0.07		0.06[b]
Married but Filed Separately	0.00	0.13		0.00	0.06		0.02[b]
Married	0.81	0.44		0.83	0.77		0.78[b]
Head of Household	0.15	0.16		0.15	0.10		0.15[b]
Tax Preparer:							
Was Tax Preparer Used?	0.60	0.67		0.67	0.58		0.62[b]
Adjusted Gross Income:	$8,042	$23,622		$8,471	$31,255		$4,234
Line 7, Wages and Salary:	$7,545	$20,766		$6,352	$29,148		$3,808
AGI income in EITC Range:							
Phase-In	0.65	0.28		0.66	0.09		0.65[b]
Flat	0.10	0.08		0.09	0.02		0.10[b]
Phase-Out	0.25	0.18		0.23	0.08		0.24[b]
Self-Employment Income:							
Fraction who Reported Any	0.06	0.07		0.30	0.06		0.11[b]
Was Tax Preparer Used?	0.79	1.00		0.72	0.70		0.78[b]
Ave Amt	$4,259	$10,740		$5,290	$13,486		$5,864[b]
EITC Received:	$1,320			$1,249			$394

Notes: a. EITC Eligibility Criterion A used.
b. Averages for those households which filed a tax return.

Table 4c
AFDC and Tax Return Outcomes, By Status of Tax Match, Eligibility, and EITC Participation
Replenishment Sample; Tax Year 1994[a]

Characteristic	*EITC-Eligible:*			*EITC-Ineligible:*			*All Households*
	EITC Participants (i)	*EITC Non-Participants* (ii)	*Did Not Have IRS Match* (iii)	*EITC Participants* (iv)	*EITC Non-Participants* (v)	*Did Not Have IRS Match* (vi)	
AFDC-FG Cases:							
% of Sample in Category:	19.55	3.22	13.31	7.26	2.90	53.76	
Months on AFDC During Year:							
0	0.08	0.11	0.17	0.11	0.05	0.12	0.12
1 to 4	0.28	0.17	0.19	0.07	0.17	0.18	0.19
5 to 8	0.22	0.16	0.09	0.10	0.15	0.12	0.14
9 to 12	0.43	0.55	0.55	0.71	0.63	0.58	0.56
Annual AFDC Benefits Received:	$3,270	$4,147	$3,951	$4,901	$4,619	$4,544	$4,231
Average EDD Wage Earnings:	$9,680	$5,804	$4,048	$601	$3,650	$186	$2,868
Tax Filing Status:							
Single	0.08	0.39		0.12	0.06		0.12[b]
Married but Filed Separately	0.00	0.00		0.00	0.05		0.00[b]
Married	0.17	0.28		0.55	0.57		0.30[b]
Head of Household	0.75	0.33		0.34	0.32		0.58[b]
Tax Preparer:							
Was Tax Preparer Used?	0.63	0.40		0.38	0.55		0.55[b]
Adjusted Gross Income:	$10,611	$18,317		$11,872	$28,127		$4,342
Line 7, Wages and Salary:	$10,265	$17,746		$10,482	$27,644		$4,141
AGI income in EITC Range:							
Phase-In	0.34	0.40		0.32	0.10		0.32[b]
Flat	0.24	0.08		0.19	0.00		0.19[b]
Phase-Out	0.41	0.10		0.49	0.10		0.37[b]
Self-Employment Income:							
Fraction who Reported Any	0.02	0.00		0.09	0.05		0.04[b]
Was Tax Preparer Used?	0.14			0.36	0.00		0.16[b]
Ave Amt	$3,250	$0		$2,462	$1,680		$2,620[b]
EITC Received:	$1,444			$1,364			$381
AFDC-U Cases:							
% of Sample in Category:	31.64	1.59	7.94	10.48	5.33	43.02	
Months on AFDC During Year:							
0	0.10	0.00	0.20	0.13	0.07	0.14	0.13
1 to 4	0.10	0.50	0.15	0.06	0.53	0.13	0.14
5 to 8	0.13	0.20	0.28	0.13	0.20	0.09	0.13
9 to 12	0.67	0.29	0.38	0.67	0.19	0.64	0.60
Annual AFDC Benefits Received:	$4,835	$3,402	$4,148	$5,359	$2,714	$5,641	$5,046
Average EDD Wage Earnings:	$7,643	$8,265	$4,417	$3,544	$23,544	$1,422	$5,138
Tax Filing Status:							
Single	0.02	0.24		0.00	0.00		0.02[b]
Married but Filed Separately	0.00	0.29		0.00	0.04		0.01[b]
Married	0.84	0.47		0.89	0.92		0.85[b]
Head of Household	0.14	0.00		0.11	0.03		0.12[b]
Tax Preparer:							
Was Tax Preparer Used?	0.64	0.77		0.79	0.71		0.68[b]
Adjusted Gross Income:	$9,600	$25,865		$10,448	$34,087		$6,360
Line 7, Wages and Salary:	$8,677	$17,138		$8,449	$31,847		$5,601
AGI income in EITC Range:							
Phase-In	0.50	0.24		0.69	0.04		0.48[b]
Flat	0.12	0.00		0.06	0.00		0.09[b]
Phase-Out	0.37	0.31		0.24	0.03		0.30[b]
Self-Employment Income:							
Fraction who Reported Any	0.04	0.00		0.37	0.04		0.11[b]
Was Tax Preparer Used?	0.69			0.84	0.33		0.66[b]
Ave Amt	$4,781	$0		$4,488	$5,882		$4,683[b]
EITC Received:	$1,534			$1,238			$615

Notes: a. EITC Eligibility Criterion A used.
b. Averages for those households which filed a tax return.

Table 5a
Number of Children in AFDC Assistance Units vs. Number of Exemptions Claimed on Tax Returns
AFDC-FG Households

	Number of Children in AFDC Assistance Unit						Percentage of Sample for which:		
Exemptions	**1**	**2**	**3**	**4**	**5**	***Total %***	AFDC < Exmptns.	AFDC = Exmptns.	AFDC > Exmptns.
Original Sample, Tax Year 1993									
0	5.02	1.96	1.17	0.42	0.40	8.97			
1	30.22	4.00	2.35	0.95	0.12	37.64			
2	4.06	18.67	4.64	0.51	0.64	28.52			
3	1.31	2.26	9.03	0.67	0.25	13.52			
4	0.30	0.43	0.46	2.98	0.07	4.24			
5	0.07	0.19	0.47	0.84	0.94	2.51			
Total %	40.98	27.51	18.12	6.37	2.42	95.40	10.89	64.82	24.29
Original Sample, Tax Year 1994									
0	4.67	2.61	0.96	0.23	0.22	8.69			
1	18.53	3.81	1.07	0.40	0.41	24.22			
2	5.22	14.03	2.08	1.18	0.36	22.87			
3	1.95	2.41	6.15	0.49	0.14	11.14			
4	0.52	0.85	0.58	1.92	0.06	3.93			
5	0.34	0.30	0.17	0.65	0.73	2.19			
Total %	31.23	24.01	11.01	4.87	1.92	73.04	17.78	56.63	25.59
Replenishment Sample, Tax Year 1994									
0	2.34	1.43	0.48			4.25			
1	28.90	4.87		0.46		34.23			
2	10.27	11.57	3.49			25.33			
3	3.12	0.56	4.09	0.21		7.98			
4	2.39			2.61	1.46	6.46			
5	0.16		0.17			0.33			
Total %	47.18	18.43	8.23	3.28	1.46	78.58	21.21	60.04	18.76

Table 5b
Number of Children in AFDC Assistance Units vs. Number of Exemptions Claimed on Tax Returns
AFDC-U Households

	Number of Children in AFDC Assistance Unit						Percentage of Sample for which:		
Exemptions	**1**	**2**	**3**	**4**	**5**	***Total %***	AFDC < Exmptns.	AFDC = Exmptns.	AFDC > Exmptns.
Original Sample, Tax Year 1993									
0	2.08	1.11	0.87	0.54	0.22	4.82			
1	13.12	2.08	0.43	0.17	0.09	15.89			
2	3.30	26.64	2.23	1.30	0.28	33.75			
3	0.48	2.67	19.85	0.91	0.30	24.21			
4	0.13	0.26	2.47	10.43	0.78	14.07			
5			0.22	0.95	4.43	5.60			
Total %	19.11	32.76	26.07	14.30	6.10	98.34	10.66	75.73	13.62
Original Sample, Tax Year 1994									
0	1.95	1.73	0.93	0.43	0.10	5.14			
1	8.56	2.04	0.33	0.19	0.23	11.35			
2	2.45	25.58	1.50	1.28	0.16	30.97			
3	0.62	2.10	14.47	0.74	0.74	18.67			
4	0.19	0.43	1.23	7.04	0.66	9.55			
5	0.02	0.23	0.16	0.70	3.66	4.77			
Total %	13.79	32.11	18.62	10.38	5.55	80.45	10.11	73.72	16.17
Replenishment Sample, Tax Year 1994									
0	1.15	0.62		0.60		2.37			
1	16.08	0.41	0.31			16.80			
2	4.04	30.74	1.32		0.30	36.40			
3	1.97	1.11	8.89	0.50	0.31	12.78			
4		0.47	1.40	5.79	0.30	7.96			
5		0.95			2.65	3.60			
Total %	23.24	34.30	11.92	6.89	3.56	79.91	12.44	80.28	7.28

Table 6a
Comparisons of AFDC Assistance Status and Tax Return Filing Status at End of Year for Adults in AFDC-FG Households at Time of Sampling

Beginning and Ending AFDC Case Statuses	Fraction Who Filed Tax Return	Tax Return Filing Status			
		Single	*Married but Filed Separate Returns*	*Married*	*Head of Household*
Original Sample, Tax Year 1993:					
AFDC-FG → AFDC-FG (N = 4,707)[a]	0.179	0.133	0.021	0.206	0.638
AFDC-FG → Other Aid (N = 1,285)[a]	0.377	0.124	0.006	0.271	0.594
AFDC-FG → AFDC-U (N = 118)[a]	0.269	0.029	0.000	0.641	0.329
Original Sample, Tax Year 1994:					
AFDC-FG → AFDC-FG (N = 4,017)[a]	0.228	0.131	0.012	0.180	0.676
AFDC-FG → Other Aid (N = 2,250)[a]	0.430	0.178	0.018	0.275	0.529
AFDC-FG → AFDC-U (N = 1,53)[a]	0.237	0.086	0.000	0.682	0.232
Replenishment Sample, Tax Year 1994:					
AFDC-FG → AFDC-FG (N = 377)[a]	0.276	0.139	0.009	0.232	0.620
AFDC-FG → Other Aid (N = 297)[a]	0.412	0.096	0.000	0.327	0.577
AFDC-FG → AFDC-U (N = 27)[a]	0.258	0.038	0.000	0.962	0.000

Note: a. Unweighted sample sizes

.

Table 6b
Comparisons of AFDC Assistance Status and Tax Return Filing Status for Adults in AFDC-U Households at Time of Sampling

Beginning and Ending AFDC Case Statuses	Fraction Who Filed Tax Return	Tax Return Filing Status			
		Single	*Married but Filed Separate Returns*	*Married*	*Head of Household*
Original Sample, Tax Year 1993:					
AFDC-U → AFDC-FG (N = 315)[a]	0.240	0.057	0.057	0.584	0.301
AFDC-U → Other Aid (N = 999)[a]	0.447	0.048	0.030	0.761	0.157
AFDC-U → AFDC-U (N = 4,272)[a]	0.268	0.019	0.007	0.906	0.065
Original Sample, Tax Year 1994:					
AFDC-U → AFDC-FG (N = 461)[a]	0.251	0.138	0.057	0.369	0.435
AFDC-U → Other Aid (N = 1,597)[a]	0.454	0.083	0.019	0.700	0.197
AFDC-U → AFDC-U (N = 3,177)[a]	0.319	0.015	0.003	0.909	0.072
Replenishment Sample, Tax Year 1994:					
AFDC-U → AFDC-FG (N = 55)[a]	0.310	0.113	0.120	0.434	0.334
AFDC-U → Other Aid (N = 436)[a]	0.494	0.011	0.021	0.776	0.192
AFDC-U → AFDC-U (N = 387)[a]	0.503	0.019	0.000	0.942	0.039

Note: a. Unweighted sample sizes

T able 7a
Comparisons of Gross Earned Income (GEI) Reported to AFDC vs. Wage Earnings Based on EDD (UI/DI) Records
AFDC-FG Households

Household's GEI Reported to AFDC	EDD Wage Earnings												
	$0	$0 - $2.5K	$2.5 - $5K	$5 - $7.5K	$7.5 - $10K	$10K - $12.5K	$12.5K - $15K	$15K - $17.5K	$17.5K - $20K	$20K - $22.5K	$22.5K - $25K	*Totals*	
Original Sample, Tax Year 1993													
Missing	1.58	0.30	0.09	0.05	0.03							2.05	
$0	74.60	8.61	1.87	1.03	0.70	0.35	0.15	0.16	0.09	0.03		87.59	
$0 - $2.5K	2.18	3.74	0.51	0.07	0.06	0.01	0.08	0.02				6.67	Percentage of Assistance Units where:
$2.5 - $5K	0.40	0.38	1.27	0.30	0.04	0.02	0.04					2.45	AFDC GEI < EDD: 14.93
$5 - $7.5K	0.13	0.05	0.12	0.31	0.11	0.06						0.78	**AFDC GEI = EDD: 81.64**
$7.5 - $10K	0.02	0.04		0.02	0.13	0.15						0.36	AFDC GEI > EDD: 3.43
$10K - $12.5K	0.05		0.02		0.02							0.09	**AFDC GEI < EDD, for EDD > 0 AUs: 68.73**
$12.5K - $15K							0.02					0.02	
$15K - $17.5K												0.00	
$17.5K - $20K												0.00	
$20K - $22.5K												0.00	
$22.5K - $25K												0.00	
Totals	78.96	13.12	3.88	1.78	1.09	0.59	0.29	0.18	0.09	0.03	0.00	100.00	
Original Sample, Tax Year 1994													
Missing	2.12	0.54	0.18	0.11	0.13		0.05	0.01		0.05	0.01	3.20	
$0	65.89	9.97	2.07	1.38	0.84	0.37	0.11	0.01	0.12	0.15	0.05	80.96	
$0 - $2.5K	3.39	4.79	1.49	0.44	0.11	0.09	0.07		0.06			10.44	Percentage of Assistance Units where:
$2.5 - $5K	0.72	0.44	1.60	0.26	0.02	0.09	0.09	0.05				3.27	AFDC GEI < EDD: 19.26
$5 - $7.5K	0.14	0.15	0.25	0.57	0.05	0.15		0.05				1.36	**AFDC GEI = EDD: 75.30**
$7.5 - $10K	0.05			0.14	0.20	0.03						0.42	AFDC GEI > EDD: 5.44
$10K - $12.5K	0.08	0.06				0.11	0.05			0.02		0.32	**AFDC GEI < EDD, for EDD > 0 AUs: 65.88**
$12.5K - $15K							0.02					0.02	
$15K - $17.5K								0.02				0.02	
$17.5K - $20K		0.02										0.02	
$20K - $22.5K												0.00	
$22.5K - $25K												0.00	
Totals	72.39	15.97	5.59	2.90	1.35	0.84	0.39	0.14	0.18	0.22	0.06	100.00	
Replenishment Sample, Tax Year 1994													
Missing	0.90	0.24	0.19									1.33	
$0	65.65	8.41	3.31	0.29		1.50		0.15			0.29	79.60	
$0 - $2.5K	3.34	5.23	0.36	0.33	0.12	0.29						9.67	Percentage of Assistance Units where:
$2.5 - $5K	0.91	1.01	2.30									4.22	AFDC GEI < EDD: 15.68
$5 - $7.5K	1.17		0.60	0.57	0.20							2.54	**AFDC GEI = EDD: 76.68**
$7.5 - $10K	0.29				2.03							2.32	AFDC GEI > EDD: 7.64
$10K - $12.5K												0.00	**AFDC GEI < EDD, for EDD > 0 AUs: 54.97**
$12.5K - $15K						0.32						0.32	
$15K - $17.5K												0.00	
$17.5K - $20K												0.00	
$20K - $22.5K												0.00	
$22.5K - $25K												0.00	
Totals	72.26	14.89	6.76	1.19	2.35	2.11	0.00	0.15	0.00	0.00	0.29	100.00	

Table 7b
Comparisons of Gross Earned Income (GEI) Reported to AFDC vs. Wage Earnings Based on EDD (UI/DI) Records
AFDC-U Households

Household's GEI Reported to AFDC	EDD Wage Earnings												
	$0	$0 - $2.5K	$2.5 - $5K	$5 - $7.5K	$7.5 - $10K	$10K - $12.5K	$12.5K - $15K	$15K - $17.5K	$17.5K - $20K	$20K - $22.5K	$22.5K - $25K	*Totals*	
Original Sample, Tax Year 1993													
Missing	1.94	0.43	0.28	0.03	0.07	0.02		0.02				2.79	
$0	65.04	4.03	0.64	0.49	0.41	0.49	0.02	0.12	0.13	0.07		71.44	
$0 - $2.5K	5.15	4.49	1.10	0.10			0.08					10.92	Percentage of Assistance Units where:
$2.5 - $5K	3.03	1.23	6.65	0.84	0.10		0.03		0.02			11.90	AFDC GEI < EDD: 9.70
$5 - $7.5K	0.72	0.02	0.43	0.97	0.07							2.21	**AFDC GEI = EDD: 79.63**
$7.5 - $10K					0.31	0.10						0.41	AFDC GEI > EDD: 10.66
$10K - $12.5K				0.07		0.13						0.20	**AFDC GEI < EDD, for EDD > 0 AUs: 36.65**
$12.5K - $15K												0.00	
$15K - $17.5K												0.00	
$17.5K - $20K												0.00	
$20K - $22.5K												0.00	
$22.5K - $25K												0.00	
Totals	75.88	10.20	9.10	2.50	0.96	0.74	0.13	0.14	0.15	0.07	0.00	100.00	
Original Sample, Tax Year 1994													
Missing	2.31	0.74	0.28	0.04	0.02	0.04					0.02	3.45	
$0	56.66	3.85	1.16	0.52	0.40	0.30	0.40	0.16	0.16	0.14	0.12	63.87	
$0 - $2.5K	7.18	5.60	1.44	0.38	0.14	0.08						14.82	Percentage of Assistance Units where:
$2.5 - $5K	3.53	1.93	6.96	0.60		0.12	0.04	0.08			0.02	13.28	AFDC GEI < EDD: 11.80
$5 - $7.5K	0.62	0.08	0.44	1.58	0.22	0.08						3.02	**AFDC GEI = EDD: 74.06**
$7.5 - $10K				0.24	0.56	0.10						0.90	AFDC GEI > EDD: 14.14
$10K - $12.5K	0.04				0.04	0.10		0.08			0.04	0.30	**AFDC GEI < EDD, for EDD > 0 AUs: 35.84**
$12.5K - $15K							0.08					0.08	
$15K - $17.5K												0.00	
$17.5K - $20K												0.00	
$20K - $22.5K												0.00	
$22.5K - $25K												0.00	
Totals	70.34	12.20	10.28	3.36	1.38	0.82	0.52	0.32	0.16	0.14	0.20	100.00	
Replenishment Sample, Tax Year 1994													
Missing	0.39		0.24		0.09							0.72	
$0	49.19	1.60	1.81	0.52			0.32	1.02				54.46	
$0 - $2.5K	4.95	6.63	2.44									14.02	Percentage of Assistance Units where:
$2.5 - $5K	6.37	3.59	5.23	1.66			1.02					17.87	AFDC GEI < EDD: 14.91
$5 - $7.5K	2.20	1.42	0.15	0.71	1.14							5.62	**AFDC GEI = EDD: 64.75**
$7.5 - $10K	1.26		0.41		1.14	2.41	0.32					5.54	AFDC GEI > EDD: 20.35
$10K - $12.5K						1.47	0.32					1.79	**AFDC GEI < EDD, for EDD > 0 AUs: 40.91**
$12.5K - $15K												0.00	
$15K - $17.5K												0.00	
$17.5K - $20K												0.00	
$20K - $22.5K												0.00	
$22.5K - $25K												0.00	
Totals	64.36	13.24	10.28	2.89	2.37	3.88	1.98	1.02	0.00	0.00	0.00	100.00	

Table 8a
Comparisons of Adjusted Gross Income (AGI) on Federal Tax Returns vs. Wage Earnings Based on EDD (UI/DI) Records
AFDC-FG Households

Amount Reported as AGI on Tax Returns	EDD Wage Earnings												
	$0	$0 - $2.5K	$2.5 - $5K	$5 - $7.5K	$7.5 - $10K	$10K - $12.5K	$12.5K - $15K	$15K - $17.5K	$17.5K - $20K	$20K - $22.5K	$22.5K - $25K	*Totals*	
Original Sample, Tax Year 1993													
No Tax Return	66.55	8.67	1.48	0.53	0.39	0.08	0.14	0.06	0.01	0.01	0.04	77.96	
$0	0.28	0.01										0.29	
$0 - $2.5K	0.76	2.27	0.14	0.01	0.01							3.19	Percentage of Assistance Units where:
$2.5 - $5K	0.71	0.66	2.45	0.15	0.06					0.01		4.04	IRS AGI < EDD: 12.61
$5 - $7.5K	1.01	0.33	0.45	1.70	0.04	0.03	0.07	0.02				3.65	**IRS AGI = EDD: 78.59**
$7.5 - $10K	0.81	0.18	0.11	0.29	1.25	0.08	0.04	0.04	0.02			2.82	IRS AGI > EDD: 8.80
$10K - $12.5K	0.42	0.12		0.09	0.20	1.09	0.01		0.01			1.94	**AGI < EDD, for EDD > 0 AUs: 3.39**
$12.5K - $15K	0.56	0.10		0.02	0.04	0.08	0.43	0.08			0.00	1.31	**EDD < AGI, for AGI > 0 AUs: 40.47**
$15K - $17.5K	0.13	0.09	0.06		0.03		0.04	0.43				0.78	
$17.5K - $20K	0.40	0.08	0.01	0.05		0.02	0.01	0.03	0.29	0.03		0.92	
$20K - $22.5K	0.23	0.01	0.02							0.18	0.08	0.52	
$22.5K - $25K	0.40	0.06	0.00						0.01		0.04	0.51	
Totals	72.26	12.58	4.72	2.84	2.02	1.38	0.74	0.66	0.34	0.23	0.16	97.93	
Original Sample, Tax Year 1994													
No Tax Return	57.35	9.08	1.92	0.91	0.45	0.25	0.11	0.07	0.07	0.04	0.10	70.35	
$0	0.04	0.06										0.10	
$0 - $2.5K	0.78	2.71	0.13	0.04	0.03	0.04				0.01		3.74	Percentage of Assistance Units where:
$2.5 - $5K	0.69	0.56	3.04	0.12	0.04	0.03	0.03					4.51	IRS AGI < EDD: 15.14
$5 - $7.5K	0.71	0.22	0.32	2.15	0.18	0.12	0.05	0.04		0.01		3.80	**IRS AGI = EDD: 74.20**
$7.5 - $10K	0.57	0.25	0.19	0.18	1.24	0.11	0.01	0.01				2.56	IRS AGI > EDD: 10.66
$10K - $12.5K	0.51	0.15	0.06	0.10	0.17	1.43	0.14	0.01		0.01	0.04	2.62	**AGI < EDD, for EDD > 0 AUs: 3.92**
$12.5K - $15K	0.67	0.25	0.07	0.06	0.09	0.07	1.04	0.05	0.01			2.31	**EDD < AGI, for AGI > 0 AUs: 36.06**
$15K - $17.5K	0.50	0.18	0.10	0.01		0.07	0.10	0.57	0.05	0.03		1.61	
$17.5K - $20K	0.61	0.18	0.01	0.07		0.04	0.06	0.11	0.59	0.02	0.01	1.70	
$20K - $22.5K	0.48	0.04	0.04	0.00	0.03	0.03	0.03	0.07	0.01	0.36	0.01	1.10	
$22.5K - $25K	0.34	0.30	0.05	0.02	0.00		0.01				0.23	0.95	
Totals	63.25	13.98	5.93	3.66	2.23	2.19	1.58	0.93	0.73	0.48	0.39	95.35	
Replenishment Sample, Tax Year 1994													
No Tax Return	53.59	8.62	1.81	1.03	0.37	0.27	0.12		0.77		0.56	67.14	
$0	0.13	0.13										0.26	
$0 - $2.5K	0.55	1.88										2.43	Percentage of Assistance Units where:
$2.5 - $5K	0.18	0.66	2.43	0.07								3.34	IRS AGI < EDD: 15.69
$5 - $7.5K	1.58	0.29	0.49	1.02	0.03							3.41	**IRS AGI = EDD: 71.31**
$7.5 - $10K	1.07	0.09		0.84	2.65	0.89						5.54	IRS AGI > EDD: 13.00
$10K - $12.5K	0.99			0.35	0.23	2.15					0.22	3.94	**AGI < EDD, for EDD > 0 AUs: 4.12**
$12.5K - $15K	0.31		0.02	0.32	0.09	0.15	0.86	0.19				1.94	**EDD < AGI, for AGI > 0 AUs: 39.88**
$15K - $17.5K	1.31	0.05		0.06			0.57	1.74	0.06			3.79	
$17.5K - $20K	1.02	0.05						0.08	0.76			1.91	
$20K - $22.5K	0.60	0.17	0.04						0.12	1.02		1.95	
$22.5K - $25K	0.08		0.06							0.13	0.60	0.87	
Totals	61.41	11.94	4.85	3.69	3.37	3.46	1.55	2.01	1.71	1.15	1.38	96.52	

Table 8b
Comparisons of Adjusted Gross Income (AGI) on Federal Tax Returns vs. Wage Earnings Based on EDD (UI/DI) Records
AFDC-U Households

Amount Reported as AGI on Tax Returns	EDD Wage Earnings												
	$0	$0 - $2.5K	$2.5 - $5K	$5 - $7.5K	$7.5 - $10K	$10K - $12.5K	$12.5K - $15K	$15K - $17.5K	$17.5K - $20K	$20K - $22.5K	$22.5K - $25K	*Totals*	
Original Sample, Tax Year 1993													
No Tax Return	61.96	5.93	1.07	0.40	0.18	0.18	0.09	0.22	0.01	0.08	0.05	70.17	
$0	0.08	0.01		0.01								0.10	
$0 - $2.5K	1.98	3.00	0.01	0.03	0.03	0.01			0.03			5.09	Percentage of Assistance Units where:
$2.5 - $5K	2.11	0.97	6.60	0.10	0.01	0.01			0.01			9.81	IRS AGI < EDD: 9.26
$5 - $7.5K	0.84	0.16	0.51	2.20	0.05	0.01	0.05	0.03				3.85	**IRS AGI = EDD: 78.39**
$7.5 - $10K	0.45	0.25	0.34	0.44	1.23	0.16			0.05			2.92	IRS AGI > EDD: 12.05
$10K - $12.5K	0.78	0.12	0.14	0.19	0.32	1.29	0.14	0.03	0.05		0.05	3.11	**AGI < EDD, for EDD > 0 AUs: 3.97**
$12.5K - $15K	0.22	0.06	0.08	0.10	0.03	0.21	0.36	0.03	0.03			1.12	**EDD < AGI, for AGI > 0 AUs: 40.53**
$15K - $17.5K	0.09			0.14		0.05	0.12	0.14	0.01	0.03	0.12	0.70	
$17.5K - $20K	0.30		0.06		0.13		0.03	0.10	0.30	0.09		1.01	
$20K - $22.5K	0.12					0.01	0.03	0.03	0.12	0.19	0.04	0.54	
$22.5K - $25K	0.10			0.05	0.03			0.04		0.05	0.08	0.35	
Totals	69.03	10.50	8.81	3.66	2.01	1.93	0.82	0.62	0.61	0.44	0.34	98.77	
Original Sample, Tax Year 1994													
No Tax Return	54.17	6.30	1.76	0.71	0.38	0.35	0.18	0.03	0.11	0.18	0.28	64.45	
$0	0.10											0.10	
$0 - $2.5K	1.71	3.07	0.11	0.07	0.03				0.03			5.02	Percentage of Assistance Units where:
$2.5 - $5K	2.22	1.10	6.20	0.22	0.08		0.03				0.01	9.86	IRS AGI < EDD: 12.31
$5 - $7.5K	0.97	0.56	0.60	2.20	0.05	0.11	0.01	0.03			0.01	4.54	**IRS AGI = EDD: 74.03**
$7.5 - $10K	0.80	0.25	0.23	0.36	1.35	0.05	0.04				0.01	3.09	IRS AGI > EDD: 13.66
$10K - $12.5K	0.58	0.10	0.18	0.25	0.37	1.08	0.15	0.03	0.03	0.03	0.03	2.83	**AGI < EDD, for EDD > 0 AUs: 4.02**
$12.5K - $15K	0.53	0.01	0.04	0.05	0.16	0.21	0.82	0.05		0.03	0.05	1.95	**EDD < AGI, for AGI > 0 AUs: 38.53**
$15K - $17.5K	0.18	0.01			0.12	0.08	0.12	0.71	0.05		0.01	1.28	
$17.5K - $20K	0.16	0.05	0.18	0.05	0.03	0.05		0.18	0.53	0.11		1.34	
$20K - $22.5K	0.14	0.05							0.11	0.43	0.08	0.81	
$22.5K - $25K	0.17	0.03	0.03				0.03	0.05		0.01	0.40	0.72	
Totals	61.73	11.53	9.33	3.91	2.57	1.93	1.38	1.08	0.86	0.79	0.88	95.99	
Replenishment Sample, Tax Year 1994													
No Tax Return	41.78	4.85	1.16	0.66	0.14	0.15	0.26	0.18	0.19		0.35	49.72	
$0	0.07											0.07	
$0 - $2.5K	1.84	1.93	0.22									3.99	Percentage of Assistance Units where:
$2.5 - $5K	2.82	1.96	4.25	0.35	0.19		0.13					9.70	IRS AGI < EDD: 11.77
$5 - $7.5K	1.98	1.44	1.56	2.27	0.31		0.51					8.07	**IRS AGI = EDD: 64.71**
$7.5 - $10K	1.04	0.07	0.48	0.14	1.38		0.13		0.07			3.31	IRS AGI > EDD: 23.51
$10K - $12.5K	0.33	0.18	0.15	0.75	0.44	3.31	0.14		0.07		0.15	5.52	**AGI < EDD, for EDD > 0 AUs: 6.12**
$12.5K - $15K	0.82	0.40	0.15	0.58	0.02	0.87	1.55	0.29				4.68	**EDD < AGI, for AGI > 0 AUs: 46.83**
$15K - $17.5K	0.30	0.35		0.26	0.07			1.56	0.02	0.07		2.63	
$17.5K - $20K	0.22	0.17		0.07	0.17	0.08		0.03	1.21			1.95	
$20K - $22.5K	0.19	0.69				0.21			0.29	0.42	0.29	2.09	
$22.5K - $25K	0.54									0.07	0.07	0.68	
Totals	51.93	12.04	7.97	5.08	2.72	4.62	2.72	2.06	1.85	0.56	0.86	92.41	

Table 9a
Comparisons of Taxable Wage and Salary Income (Line 7 Income on Tax Returns) vs. Wage Earnings Based on EDD (UI/DI) Records
AFDC-FG Households

Amount Reported on Line 7 of Tax Returns	EDD Wage Earnings												
	$0	$0 - $2.5K	$2.5 - $5K	$5 - $7.5K	$7.5 - $10K	$10K - $12.5K	$12.5K - $15K	$15K - $17.5K	$17.5K - $20K	$20K - $22.5K	$22.5K - $25K	*Totals*	
Original Sample, Tax Year 1993													
No Tax Return	66.48	8.67	1.48	0.53	0.39	0.08	0.14	0.06	0.01	0.01	0.04	77.89	
$0	1.22	0.10	0.04								0.01	1.37	
$0 - $2.5K	0.76	2.66	0.14	0.01	0.01				0.01			3.59	Percentage of Assistance Units where:
$2.5 - $5K	0.52	0.48	2.65	0.18	0.06					0.01		3.90	Line 7 W&S < EDD: 12.79
$5 - $7.5K	0.67	0.19	0.26	1.84	0.04	0.06	0.07	0.02				3.15	**Line 7 W&S = EDD: 80.23**
$7.5 - $10K	0.73	0.10	0.06	0.20	1.33	0.08	0.04	0.04	0.02			2.60	Line 7 W&S > EDD: 6.99
$10K - $12.5K	0.41	0.06		0.02	0.13	1.17	0.01		0.01		0.00	1.81	**AGI < EDD, for EDD > 0 AUs: 4.13**
$12.5K - $15K	0.45	0.13		0.01	0.04	0.04	0.46	0.08				1.21	**EDD < Line 7, for Line 7 > 0 AUs: 33.70**
$15K - $17.5K	0.16	0.06	0.07		0.03		0.01	0.43				0.76	
$17.5K - $20K	0.39	0.09	0.00	0.05		0.01	0.01	0.03	0.28	0.03		0.89	
$20K - $22.5K	0.26	0.02	0.02							0.18	0.07	0.55	
$22.5K - $25K	0.32	0.06	0.00						0.01		0.05	0.44	
Totals	72.37	12.62	4.72	2.84	2.03	1.44	0.74	0.66	0.34	0.23	0.17	98.16	
Original Sample, Tax Year 1994													
No Tax Return	57.31	9.04	1.92	0.91	0.45	0.25	0.11	0.07	0.07	0.04	0.10	70.27	
$0	0.81	0.12	0.04	0.03								1.00	
$0 - $2.5K	0.78	3.01	0.15	0.04	0.01	0.04				0.01		4.04	Percentage of Assistance Units where:
$2.5 - $5K	0.53	0.43	3.27	0.12	0.07	0.03	0.03					4.48	Line 7 W&S < EDD: 13.52
$5 - $7.5K	0.60	0.23	0.15	2.21	0.19	0.12	0.07	0.04		0.01		3.62	**Line 7 W&S = EDD: 77.25**
$7.5 - $10K	0.31	0.15	0.09	0.13	1.30	0.16		0.01				2.15	Line 7 W&S > EDD: 9.23
$10K - $12.5K	0.54	0.15	0.08	0.07	0.09	1.45	0.17	0.00		0.01	0.04	2.60	**AGI < EDD, for EDD > 0 AUs: 4.69**
$12.5K - $15K	0.69	0.27	0.04	0.06	0.09	0.03	1.05	0.07	0.01			2.31	**EDD < Line 7, for Line 7 > 0 AUs: 30.00**
$15K - $17.5K	0.43	0.15	0.08	0.01		0.07	0.10	0.65	0.06	0.03		1.58	
$17.5K - $20K	0.58	0.15	0.02	0.08		0.04	0.06	0.03	0.58	0.02	0.01	1.57	
$20K - $22.5K	0.53	0.08	0.03	0.03	0.03	0.00		0.07	0.01	0.36	0.01	1.15	
$22.5K - $25K	0.21	0.27	0.05	0.01	0.00						0.23	0.77	
Totals	63.32	14.05	5.92	3.70	2.23	2.19	1.59	0.94	0.73	0.48	0.39	95.54	
Replenishment Sample, Tax Year 1994													
No Tax Return	53.59	8.62	1.81	1.03	0.37	0.27	0.12		0.77		0.56	67.14	
$0	0.88	0.05										0.93	
$0 - $2.5K	0.47	2.26	0.13									2.86	Percentage of Assistance Units where:
$2.5 - $5K	0.03	0.49	2.57	0.07	0.01							3.17	Line 7 W&S < EDD: 15.62
$5 - $7.5K	1.22	0.29	0.23	1.96	0.01							3.71	**Line 7 W&S = EDD: 73.49**
$7.5 - $10K	1.23			0.18	2.68	0.89						4.98	Line 7 W&S > EDD: 10.89
$10K - $12.5K	0.68			0.06	0.21	2.15					0.22	3.32	**AGI < EDD, for EDD > 0 AUs: 4.28**
$12.5K - $15K	1.31		0.02	0.32	0.09	0.15	0.86	0.19				2.94	**EDD < Line 7, for Line 7 > 0 AUs: 34.10**
$15K - $17.5K	0.88	0.05		0.06			0.57	1.82	0.06			3.44	
$17.5K - $20K	1.02								0.88			1.90	
$20K - $22.5K		0.17	0.04							1.15		1.36	
$22.5K - $25K	0.62		0.06				0.13				0.60	1.41	
Totals	61.93	11.93	4.86	3.68	3.37	3.46	1.68	2.01	1.71	1.15	1.38	97.16	

Table 9b
Comparisons of Taxable Wage and Salary Income (Line 7 Income on Tax Returns) vs. Wage Earnings Based on EDD (UI/DI) Records
AFDC-U Households

Amount Reported on Line 7 of Tax Returns	EDD Wage Earnings												
	$0	$0 - $2.5K	$2.5 - $5K	$5 - $7.5K	$7.5 - $10K	$10K - $12.5K	$12.5K - $15K	$15K - $17.5K	$17.5K - $20K	$20K - $22.5K	$22.5K - $25K	*Totals*	
Original Sample, Tax Year 1993													
No Tax Return	61.92	5.93	1.07	0.40	0.18	0.18	0.09	0.22	0.01	0.08	0.05	70.13	
$0	2.42	0.17	0.12						0.01			2.72	
$0 - $2.5K	1.42	3.28	0.05	0.04	0.06	0.01			0.03			4.89	Percentage of Assistance Units where:
$2.5 - $5K	1.49	0.83	7.08	0.13		0.01						9.54	Line 7 W&S < EDD: 9.87
$5 - $7.5K	0.30	0.10	0.20	2.69	0.05	0.08	0.05	0.03				3.50	**Line 7 W&S = EDD: 82.85**
$7.5 - $10K	0.35	0.13	0.21	0.13	1.44	0.10			0.06			2.42	Line 7 W&S > EDD: 7.28
$10K - $12.5K	0.47	0.05	0.09	0.09	0.19	1.48	0.09	0.03	0.04		0.05	2.58	**AGI < EDD, for EDD > 0 AUs: 4.93**
$12.5K - $15K	0.08	0.01		0.05		0.06	0.52	0.03				0.75	**EDD < Line 7, for Line 7 > 0 AUs: 26.83**
$15K - $17.5K	0.09			0.09			0.06	0.21	0.01	0.03	0.12	0.61	
$17.5K - $20K	0.26				0.08			0.06	0.43	0.09		0.92	
$20K - $22.5K	0.14					0.01			0.01	0.22	0.04	0.42	
$22.5K - $25K	0.05			0.05				0.01		0.03	0.10	0.24	
Totals	68.99	10.50	8.82	3.67	2.00	1.93	0.81	0.59	0.60	0.45	0.36	98.72	
Original Sample, Tax Year 1994													
No Tax Return	54.11	6.30	1.76	0.71	0.38	0.35	0.18	0.03	0.11	0.18	0.28	64.39	
$0	2.72	0.15										2.87	
$0 - $2.5K	1.39	3.31	0.16	0.11	0.03				0.03			5.03	Percentage of Assistance Units where:
$2.5 - $5K	1.34	1.04	6.64	0.21	0.08		0.03				0.01	9.35	Line 7 W&S < EDD: 12.81
$5 - $7.5K	0.61	0.47	0.38	2.49	0.12	0.11	0.01	0.03			0.01	4.23	**Line 7 W&S = EDD: 77.97**
$7.5 - $10K	0.27	0.12	0.10	0.19	1.63	0.05	0.04			0.03	0.01	2.44	Line 7 W&S > EDD: 9.21
$10K - $12.5K	0.44	0.03	0.08	0.18	0.29	1.15	0.32	0.08	0.03		0.05	2.65	**AGI < EDD, for EDD > 0 AUs: 5.42**
$12.5K - $15K	0.41	0.01	0.01		0.07	0.15	0.73	0.07		0.03	0.03	1.51	**EDD < Line 7, for Line 7 > 0 AUs: 28.14**
$15K - $17.5K	0.16	0.01	0.05		0.01	0.12	0.05	0.84	0.05	0.05	0.01	1.35	
$17.5K - $20K	0.08	0.05	0.11	0.05	0.03			0.04	0.62	0.05		1.03	
$20K - $22.5K	0.16	0.05					0.03		0.04	0.43	0.08	0.79	
$22.5K - $25K	0.14	0.03	0.03					0.05		0.01	0.49	0.75	
Totals	61.83	11.57	9.32	3.94	2.64	1.93	1.39	1.14	0.88	0.78	0.97	96.39	
Replenishment Sample, Tax Year 1994													
No Tax Return	41.63	4.85	1.16	0.66	0.14	0.15	0.26	0.18	0.19		0.35	49.57	
$0	3.82											3.82	
$0 - $2.5K	1.39	3.30	0.29									4.98	Percentage of Assistance Units where:
$2.5 - $5K	2.19	2.02	5.25	0.35	0.19		0.13			0.07		10.20	Line 7 W&S < EDD: 11.06
$5 - $7.5K	0.36	0.91	0.63	3.68	0.31							5.89	**Line 7 W&S = EDD: 73.85**
$7.5 - $10K	0.52		0.48	0.06	2.02		0.13		0.07			3.28	Line 7 W&S > EDD: 15.08
$10K - $12.5K	0.16	0.18			0.07	4.02	0.14		0.07		0.15	4.79	**AGI < EDD, for EDD > 0 AUs: 4.92**
$12.5K - $15K	0.82	0.40	0.15	0.51		0.18	2.06	0.15				4.27	**EDD < Line 7, for Line 7 > 0 AUs: 32.36**
$15K - $17.5K	0.15	0.21		0.26		0.28		1.27	0.02			2.19	
$17.5K - $20K	0.31	0.17		0.07				0.47	1.21			2.23	
$20K - $22.5K	0.19								0.29	0.42	0.29	1.19	
$22.5K - $25K	0.54									0.07	0.07	0.68	
Totals	52.08	12.04	7.96	5.59	2.73	4.63	2.72	2.07	1.85	0.56	0.86	93.09	

APPENDIX
Further Comparison of the Sample Characteristics of the Full CWAD Sample and the Subsample Used in our Analysis

In this Appendix, we briefly compare the sample characteristics of the entire sample of assistance units used in the California Work Pays Demonstration Project (CWPDP)—i.e., the sample that comprises the County Welfare Administrative Database (CWAD)—with the subsample we used in our analysis of EITC eligibility, participation and compliance rates. See Section 2.1 for a discussion of the selections we performed on the full CWAD sample to reach our analysis sample.

Characteristics of Full CWAD Sample:

Table A1 shows characteristics of the CWAD sample, by county and for the total sample, for the original (October 1992) sample.[37] Table A2 shows similar information for the Replenishment Sample. The Original Sample consists of 22,074 persons in 8,316 AFDC-FG cases and 16,005 persons in 3,975 AFDC-U cases. The Replenishment Sample contains 1,914 persons in 1,005 AFDC-FG cases and 2,223 persons in 845 AFDC-U cases. Most AFDC-FG cases in each sample have one adult and one or two children, while most AFDC-U cases in the Original Sample have two adults and two or more children. Child-only cases make up just over 30 percent of AFDC-FG cases and 23 percent of AFDC-U cases in the Original Sample, but almost 40 percent of AFDC-FG and AFDC-U cases in the Replenishment Sample.[38]

In the Original Sample, about 41 percent of persons in AFDC-FG cases are Hispanic, 34 percent black, 19 percent white, and 6 percent Asian. Hispanics and whites constitute somewhat higher percentages of AFDC-FG cases in the Replenishment Sample. Persons in AFDC-U cases from the Original Sample are 33 percent Hispanic, 6 percent black, 35 percent white, and 26 percent Asian. Over 55 percent of the AFDC-U Replenishment Sample are Hispanic. In addition to racial composition differences among the original and replacement samples across AFDC-FG and AFDC-U cases, racial composition varies among the four counties and within counties by type of AFDC case. For example, 60 percent of persons in the Alameda AFDC-FG Original Sample are black, and over 40 percent of persons in the San Bernardino AFDC-FG original and Replenishment Samples are white.

Base Wage File matches were found for just over half of the adults, though about 80 percent of adults in AFDC-FG cases and 70 to 75 percent of adults in AFDC-U cases have zero earnings recorded in the Base Wage File during the year.[39]

Characteristics of Our Analysis Sample:

Table 2a shows characteristics of the Original Sample, by county and for the total sample,

[37] The descriptive statistics reported here reflect case composition at the time of sampling, and do not reflect case composition over time. Additionally, we dropped treatment cases in San Joaquin county prior to examining baseline statistics, for reasons described in Section 2.1 of this report.

[38] Percentages reported in this Appendix are based on weighted sample data.

[39] The year refers to 1993 for Original Sample members, and 1994 for Replenishment Sample.

after we impose the restrictions described in Section 2.1. Table 2b shows similar information for the Replenishment Sample. The restricted Original Sample now has 6,356 adults in 6,036 AFDC-FG cases and 5,651 adults in 3,045 AFDC-U cases.[40] The Replenishment Sample contains 709 adults in 670 AFDC-FG cases and 890 adults in 552 AFDC-U cases. All AFDC-FG and AFDC-U cases have either one or two adults, due to the sample restrictions, with one adult in over 93 percent of AFDC-FG cases and two adults in 83 percent of the AFDC-U Original Sample and just over 51 percent of the AFDC-U Replenishment Sample.

The effects of our sample restrictions on racial composition generally seem to reduce the percentage of Hispanics across samples. This result may be expected due to our screening of child-only cases, in which Hispanics may be represented disproportionately. In the restricted Original Sample, about 29 percent of persons in AFDC-FG cases are Hispanic, 39 percent Black, 24 percent white, and 7 percent Asian. Restricted Replenishment Sample AFDC-FG cases consist of 38 percent Hispanic, 21 percent black, 35 percent white, and 4 percent Asian. Persons in AFDC-U cases from the restricted Original Sample are 22 percent Hispanic, 6 percent black, 46 percent white, and 25 percent Asian. Hispanics now account for about 47 percent of the restricted AFDC-U Replenishment Sample.

Matches to Base Wage File records were located for 76 percent of AFDC-FG adults and about 70 percent of AFDC-U adults in the restricted Original Sample, and for over 80 percent of adults in the restricted Replenishment Sample. About 70 percent of adults in the restricted Original Sample have zero earnings recorded in the Base Wage File during the year. Just over 60 percent of AFDC-FG adults and 50 percent of AFDC-U adults in the restricted Replenishment Sample show zero earnings during the year.

[40] Numbers of cases in the analysis sample do not necessarily reflect percentage reductions for child-only and other cases from the number of cases in the CWAD baseline sample because percentages are based on weighted sample data.

Table A1
Characteristics of Full CWPDP Sample, before Exclusions
Original Sample, Tax Year 1993

Variables	County				
	Alameda	*Los Angeles*	*San Bernardino*	*San Joaquin*	*Four Counties*
AFDC-FG Cases					
Race:					
White	0.15	0.13	0.41	0.35	0.19
Black	0.60	0.33	0.22	0.12	0.34
Hispanic	0.12	0.47	0.34	0.24	0.41
Asian	0.11	0.06	0.02	0.28	0.06
Gender: (fraction male)	0.37	0.40	0.39	0.40	0.39
Number of Adults in AU:					
0	0.18	0.38	0.19	0.18	0.33
1	0.80	0.58	0.76	0.77	0.63
2	0.02	0.04	0.05	0.04	0.04
3+	0.00	0.00	0.00	0.00	0.00
Number of Children in AU:					
0	0.03	0.01	0.02	0.01	0.01
1	0.44	0.40	0.35	0.35	0.39
2	0.29	0.31	0.33	0.31	0.31
3+	0.23	0.29	0.30	0.33	0.28
No. of Adults with EDD Earnings:					
0	0.75	0.84	0.78	0.76	0.82
1	0.24	0.16	0.21	0.23	0.18
2	0.01	0.00	0.01	0.01	0.00
No. of Assistance Units	1,942	3,762	1,929	683	8,316
AFDC-U Cases					
Race:					
White	0.26	0.34	0.45	0.22	0.35
Black	0.11	0.04	0.10	0.03	0.06
Hispanic	0.09	0.37	0.34	0.13	0.33
Asian	0.53	0.24	0.11	0.62	0.25
Gender: (Fraction Male)	0.50	0.51	0.49	0.51	0.51
Number of Adults in AU:					
0	0.06	0.28	0.14	0.05	0.23
1	0.07	0.12	0.13	0.06	0.12
2	0.81	0.55	0.70	0.81	0.60
3+	0.05	0.05	0.03	0.08	0.05
Number of Children in AU:					
0	0.04	0.01	0.02	0.00	0.01
1	0.16	0.21	0.16	0.14	0.20
2	0.26	0.38	0.29	0.23	0.35
3+	0.55	0.40	0.53	0.62	0.44
No. of Adults with EDD Earnings:					
0	0.76	0.77	0.66	0.63	0.75
1	0.20	0.20	0.26	0.30	0.21
2	0.04	0.03	0.08	0.06	0.04
No. of Assistance Units	849	1,951	922	253	3,975

Table A2
Characteristics of Full CWPDP Sample, before Exclusions
Replenishment Sample, Tax Year 1994

Variables	County				
	Alameda	*Los Angeles*	*San Bernardino*	*San Joaquin*	*Four Counties*
AFDC-FG Cases					
Race:					
White	0.25	0.24	0.47	0.47	0.30
Black	0.43	0.16	0.18	0.14	0.18
Hispanic	0.21	0.54	0.33	0.30	0.46
Asian	0.10	0.02	0.01	0.07	0.03
Gender: (fraction male)	0.29	0.35	0.36	0.34	0.35
Number of Adults in AU:					
0	0.27	0.44	0.28	0.24	0.39
1	0.68	0.55	0.68	0.72	0.59
2	0.05	0.01	0.04	0.04	0.02
3+	0.00	0.00	0.00	0.00	0.00
Number of Children in AU:					
0	0.25	0.28	0.29	0.43	0.28
1	0.41	0.43	0.36	0.31	0.41
2	0.23	0.18	0.20	0.13	0.18
3+	0.11	0.12	0.14	0.12	0.12
No. of Adults with EDD Earnings:					
0	0.64	0.81	0.71	0.62	0.78
1	0.34	0.19	0.29	0.36	0.22
2	0.02	0.00	0.00	0.02	0.00
No. of Assistance Units	187	420	339	59	1,005
AFDC-U Cases					
Race:					
White	0.28	0.21	0.45	0.32	0.27
Black	0.13	0.05	0.06	0.00	0.05
Hispanic	0.30	0.63	0.43	0.56	0.56
Asian	0.29	0.11	0.05	0.11	0.10
Gender: (Fraction Male)	0.49	0.52	0.51	0.44	0.52
Number of Adults in AU:					
0	0.23	0.43	0.21	0.10	0.37
1	0.22	0.29	0.31	0.22	0.29
2	0.52	0.25	0.47	0.64	0.31
3+	0.02	0.03	0.01	0.03	0.03
Number of Children in AU:					
0	0.22	0.25	0.24	0.56	0.25
1	0.27	0.26	0.22	0.05	0.25
2	0.38	0.30	0.26	0.15	0.29
3+	0.13	0.19	0.29	0.25	0.21
No. of Adults with EDD Earnings:					
0	0.65	0.76	0.51	0.27	0.70
1	0.28	0.20	0.37	0.46	0.24
2	0.07	0.04	0.12	0.24	0.06
No. of Assistance Units	112	403	287	43	845

Examining the Effect of the Earned Income Tax Credit on the Labor Market Participation of Families on Welfare*

V. Joseph Hotz
Department of Economics and NBER
University of California, Los Angeles
hotz@ucla.edu

Charles H. Mullin
Bates White
charlie.mullin@bateswhite.com

and

John Karl Scholz
Department of Economics, Institute for Research on Poverty and NBER
University of Wisconsin - Madison
jkscholz@facstaff.wisc.edu

Latest Draft: September 25, 2005

*We thank Liz Davis, Janet Holtzblatt, Tom MaCurdy, Bruce Meyer, and numerous seminar participants for comments, and Dawn Duren, Eduardo Fajnzylber, Webb Hester, Doug Houston, Jacob Klerman, Doug McKee, Oscar Mitnik, Don Oellerich, Juan Pantano, George Ramsey, Pat Ruggles and Werner Shenk for their help with this project. We also thank colleagues at the California Department of Social Services; the California Franchise Tax Board; RAND; the National Science Foundation; and the Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation who collectively helped support this work. None of these people or organizations, however, are responsible for the views expressed in this paper.

1. Introduction

Between 1990 and 1999, real spending on the earned income tax credit (EITC) increased to $31.9 billion from $9.6 billion (in 1999 dollars). It is by far the largest cash or near-cash antipoverty policy in the United States. Moreover, with phased-in increases enacted in 1990 and in 1993, it became the most rapidly growing (substantial) item in the federal budget.

Employment rates of single women with children also rose sharply over this period and welfare caseloads fell precipitously. Specifically, between March 1990 and March 2000, employment rates of single women with children rose to 73.9 percent from 55.2 percent. Welfare caseloads fell to 2.3 million families from 4.1 million (though caseloads rose until 1994, when they peaked at 5.0 million families).[1]

The coincident timing of these trends raises a question: did the EITC play a substantial role in the increase in labor force participation of single women with children?[2] Not surprisingly, given the important role the EITC plays in the nation's safety net, a number of previous studies examine this topic.[3] These papers come to a consistent conclusion: the EITC has a significant, empirically large effect on labor force participation of single women with children.[4]

---

[1] Employment rates and caseload numbers come from Tables No. 577 and No. 544, respectively, from *Statistical Abstract of the United States 2000*.

[2] Standard labor-leisure models unambiguously predict that EITC increases (or increases in other wage subsidies) will increase employment among low-skilled workers.

[3] For published work, see Dickert, Houser and Scholz (1995); Eissa and Leibman (1996); Keane and Moffitt (1998); Ellwood (2000); Meyer and Rosenbaum (2000, 2001); Grogger (2003); and Eissa and Hoynes (2004). Hotz and Scholz (2003) survey EITC research.

[4] Hotz and Scholz (2003) compute employment elasticities with respect to net income (associated with EITC changes) from selected previous studies that range from 0.69 to 1.16. Grogger's (2003) estimates were not included in our earlier study, but he concludes that the EITC "may be the single most important policy measure for explaining the decrease in welfare and the rise in work and earnings among female-headed families in recent years" (p. 408). Eissa and Hoynes (2004) focus on the employment and hours decisions of secondary workers in married families and find small, negative effects of the credit on work. Cancian and Levinson (2003) use data from the National Survey of American Families and focus on Wisconsin's supplemental EITC for families with three or more children. Unlike other studies on the topic, they do not find statistically significant EITC employment effects.

At least three concerns arise with previous work examining the EITC's employment effects. The identification strategy in most papers is to compare employment changes of groups eligible for the EITC (generally single mothers) with employment changes of groups ineligible for the EITC (generally single women without children) before and after EITC increases.[5] But other factors change over the periods the EITC has been increased, and these other factors might have different employment effects for women with and without children. Many states, for example, made extensive changes to welfare both before and after the Personal Responsibility, Welfare and Opportunity Reconciliation Act (PRWORA) abolished AFDC in 1996. And aggregate GDP increased 120 consecutive months beginning in March 1991, resulting in the longest economic expansion in U.S. history.

A second issue arises with the difference-in-difference approach: namely that the composition of the groups may change over time. There are *a priori* reasons to be concerned about potential compositional bias in these repeated cross-sectional studies.[6] Between 1984 and 1996, for example, the period examined in Meyer and Rosenbaum (2001), the fraction of all families that were single mothers in the CPS increased 30.5 percent, to 7.7 million families from 5.9 million families.[7] The difference-in-difference identification strategy will be biased in indeterminate directions if employment propensities of the sample changed over time after

[5] Keane and Moffitt (1998) estimate utility function parameters of households that are consistent with participation decisions in multiple welfare programs, accounting for the precise budget sets that occur in the 1984 SIPP. Dickert, Houser, and Scholz (1995) do not observe EITC variation in their cross-sectional study and consequently they infer the potential effects of the EITC from observing the correlation between employment and differences in the return to work across states. These employment differences, however, may be due to unobserved state-specific factors unrelated to differences in the return to work generated by the tax and transfer systems.

[6] All previous studies use repeated cross-sectional data from the Current Population Survey (CPS) and the Survey of Income and Program Participation (SIPP) except for Cancian and Levinson (2003).

[7] As discussed in Hotz and Scholz (2002), the ratio of AFDC recipients reported in the CPS to administrative counts of recipients fell from 86.7 percent in 1990 to 79.6 percent in 1996. The ratio of AFDC dollars to administrative totals was 78.4 percent in the 1984 CPS and 67.7 percent by 1996. By all accounts the SIPP and CPS are high quality, but these trends raise a concern about the ability of studies using the CPS or SIPP to accurately characterize the tax and transfer environment facing low-income families.

accounting for observed characteristics (Heckman, 1996).

Third, all previous EITC employment studies have been *indirect* in the sense that they have inferred the EITC's effects based on the relationship between employment and EITC policy changes. But if EITC changes cause employment changes, we should see similar changes in EITC claims. Because of data limitations, no previous study has examined directly EITC claiming and employment.

This paper addresses these and other concerns. We use data only from California to mitigate the influence of changes in welfare and other social policies on employment, and to better account for local labor market conditions.[8] Using data with county identifiers we are able to include considerably more detailed covariates to proxy for local labor market effects than have previous studies.[9] Moreover, California also made relatively modest changes to AFDC and TANF compared to other states in the 1990s. We nevertheless also condition on several features of the county-administered welfare programs in California.

Ours is also the first EITC employment study to use longitudinal data, which allow us to account for unobserved, household-specific, time-invariant factors that affect employment. Longitudinal data also help us avoid bias due to (unobserved) changes in the composition of "treatment" and "comparison" groups in repeated cross sectional studies.

We use a different identification strategy than most previous studies, which compare women with children to others without. Beginning in 1991, the maximum EITC available to families with two or more children increased relative to the EITC available for one-child families. While

[8] See Ellwood (2000) for more on these issues.

[9] Bartik and Eberts (1999) criticize studies that include only unemployment rates to account for labor market conditions, writing, the "… unemployment rate by itself may be a woefully incomplete measure of economic conditions affecting potential welfare recipients." They argue that more textured measures of the economic environment facing welfare families matter in understanding caseload changes.

the difference in the maximum credit by number of children was initially modest, starting in 1994, as a result of changes in the EITC included in the 1993 federal budget act, this differential was substantially increased, with the maximum credit going from $77 in 1993 to $490 in 1993 to $1,404 in 1996 (see Table 1). If the EITC alters employment, all else being equal, employment rates for two-or-more child families should grow relative to the employment rates of one-child families, as credit amounts available to these groups of families diverge.[10] Our empirical approach formalizes this strategy as a "difference-in-difference" estimator to identify the effects of the EITC on employment and other outcomes.

While the above intuition for identifying the effects of the EITC may seem compelling, it is subject to a potentially important criticism, namely that household fertility and employment levels are jointly determined, which can give rise to a form of omitted variable bias. To mitigate this bias, we use multivariate methods to adjust for observable factors and include household-specific fixed effects using our longitudinal data to adjust for time-invariant observed and unobserved factors. Furthermore, we conduct an independent test of the validity of unadjusted and adjusted versions of our identification strategy. In particular, if the EITC and not other factors (such as the strong economy) is causing employment differences between families with two or more children relative to those with one child, we should expect to see no employment differences (after conditioning on other characteristics) between families with two children and families with three or more children, since the EITC did not change differentially for the latter two groups.

Our work also differs from previous work by focusing on the portion of the population that

[10] In a sensitivity analysis section Meyer and Rosenbaum (2001) examine differences in labor market effects generated by the EITC for families with two or more children relative to families with one child, as is done in our paper. They find much smaller EITC effects in this alternative specification and the coefficients are insignificant in their larger, CPS Outgoing Rotation Group sample. Cancian and Levinson (2003) use a similar approach and find

is of considerable interest to policymakers: women with children who are trying to make the transition from welfare to work. Existing studies typically focus on single mothers using data from the CPS and SIPP. But labor force participation rates of single mothers in the CPS are around 80 percent. Labor force participation rates in our sample of welfare recipients start around 35 percent. Around 28 percent of single mothers in the CPS receive welfare. All households in our sample received welfare at some point. It is not clear whether previous results for single mothers also apply to the more economically disadvantaged population of welfare recipients.

Lastly, the most novel feature of our paper is that, through a unique match of administrative data on welfare receipt, unemployment records, and the federal tax returns filed by California residents, we are able to examine directly the employment and EITC-claiming behavior of low-income families.[11] No previous EITC employment study has used data on whether or not people eligible for the EITC actually file tax returns and claim it. If the employment patterns previous authors attribute to the EITC are, in fact, *caused* by the credit, those who are affected by the credit should file and claim it.

Using fixed effects empirical employment models estimated on a sample of single-parent families, our coefficient estimates are consistent with the EITC having a substantial, positive effect on the employment of families who have or will use welfare. Three supplementary analyses increase our confidence that we are identifying EITC effects. First, the differential employment responses occur only for families with two or more children relative to one child, and not for families with three or more children relative to families with exactly two children.

---

insignificant EITC employment effects.

[11] All tax data are processed at the California Franchise Tax Board, where strict procedures are followed (such as presenting data only in grouped format and removing all identifying information) to ensure compliance with tax privacy laws.

Leading alternative explanations, such as differential effects of the strong business cycle expansion on families with different numbers of children, would be unlikely to result in the employment patterns we find. Second, patterns of EITC claiming mirror the employment patterns. This strikes us as being a "necessary condition" for the EITC having a causal effect on employment. Third, there is no differential expansion of employment for families with two or more children relative to one-child families for those who do not file tax returns. Again, if factors other than the EITC were driving the employment patterns observed in the data, we see no reason why they would only affect those who file tax returns.

We also provide extensive sensitivity analyses that show our results are robust to several alternative modeling choices. The most striking aspect of the sensitivity analysis, however, is that the EITC has no discernable effect on employment for AFDC-UP (two parent) families.

2. The EITC and California Administrative Data

In 1999, taxpayers with two or more children could receive an EITC of 40 percent of income up to $9,540, for a maximum credit of $3,816. Taxpayers (with two or more children) with earnings between $9,540 and $12,460 receive the maximum credit. Their credit is reduced by 21.06 percent of earnings between $12,460 and $30,585. Table 1 shows the complete evolution of income eligibility thresholds, credit rates, and phase-out (or implicit tax) rates.

To receive the credit taxpayers file their regular tax return and fill out the six-line Schedule EIC that gathers information about qualifying children. The EITC is refundable, meaning that the Treasury pays it out regardless of whether the taxpayer has any Federal income tax liability. There are several basic tests for EITC eligibility. The taxpayer's earned income and adjusted gross income must be below a threshold that varies by year and by family size. To receive the

credit available to families with children,[12] the qualifying child must be younger than 19, younger than 24 if a full-time student, or any age if totally disabled. The claimant must be the parent, the grandparent, or foster parent of the child.[13] The qualifying child must live with the taxpayer at least six months during the year.

A key development for the purposes of this paper was put in place as part of the 1990 EITC expansions. After 1990, for the first time, families with two or more children were able to receive a larger EITC than they could if they had only one child. The difference through 1993, however, never exceeded $77. As part of the 1993 EITC expansion, the differences became much larger: in 1994 the maximum difference was $490, it was $1,016 in 1995, and $1,404 in 1996 (and indexed for inflation thereafter). As noted in the introduction, if the EITC alters employment, all else being equal, employment rates for two-or-more child families should grow relative to the employment rates of one-child families starting in 1994, as credit amounts available to these groups of families diverge.

For those out of the labor market, the EITC provides an unambiguous, positive incentive to work. We focus on this feature of the credit in this paper.[14]

2.1 The Core Administrative Data Sources

California's Medi-Cal Eligibility Data System (MEDS) provides information on program participation and demographic characteristics of participants in Medi-Cal, the California state Medicaid program, and AFDC/CalWORKS, the California state welfare program.[15] MEDS also

[12] A small credit available for childless taxpayers between the ages 24 in 65 with very low incomes was added in 1994. The credit rate for these taxpayers is 7.65 percent and the maximum credit in 1999 is $347.

[13] Until late 1999, a foster child was any child for whom the claimant cared for "as if the child is his or her own." Now the caring stipulation still holds, but the child must also be placed in the home by an authorized placement agency.

[14] See Hotz and Scholz (2003) for a summary of papers examining the effects of the EITC on hours of work.

[15] More generally, the MEDS is a statewide administrative system that contains information on monthly

provides AFDC participation histories of individuals in our samples starting in 1986.

We measure labor force participation using quarterly data on employment (and earnings) from the California Employment Development Department (EDD) Base Wage Files. The EDD Base Wage File contains employer-reported taxable wage payments for jobs covered by unemployment insurance (UI) and disability insurance (DI).[16] Hotz and Scholz (2002) survey studies that examine the accuracy and coverage of unemployment insurance data for the low-income population. In brief, UI data exclude some groups of workers, such as the self-employed, military, federal employees, and independent contractors. Nevertheless, employment rates derived from UI data appear to be similar to those that result from survey data. We expect UI-based employment rates to be lower because of coverage problems with flexible workers/independent contractors. But surveys suffer from nonresponse, so undercounts in both data sources typically appear to be similar in practice.

To examine tax filing and EITC claiming, we use data from the federal tax returns filed by households in the MEDs data between 1991 and 2000.

2.2 Sampling

Our core data come from a sample, drawn by scholars at Rand from the statewide MEDS data, of all assistance units ever on welfare in California between 1987 and 2000. The number of California welfare recipients is very large, so the sample was drawn to maintain approximately equal numbers of observations in each county (with every observation included in smaller counties). Sampling is based on the county in which recipients are first observed receiving cash

---

participation in the state's Medicaid program (Medi-Cal), AFDC/TANF programs, as well as the Food Stamps, SSI and California's General Assistance (GA) programs. MEDS data are used by the Department of Health and Human Services for quality control analyses and are thought to be high quality for the welfare and Medicaid populations.

[16]The file generally includes individuals paid cash wages of more than $100 in a calendar quarter, and domestic workers paid cash wages more than $750 in a calendar quarter.

assistance and time-invariant sampling weights are applied to each person. Approximately 50 percent of all adults ever on welfare over this period are included in the baseline Rand sample.[17] We start with the Rand sample because it was linked to employment data beginning in 1991 from the California unemployment insurance system. Statewide unemployment insurance data from early in the analysis period got destroyed sometime in the 1990s, making the Rand sample an essential starting point for the analysis.

Sampling refinements are invariably needed when working with administrative data. We drop roughly 6 percent of households in which no adult is a member of the AFDC assistance unit, since we do not have earnings and demographic information on the potential EITC claimant.[18] We drop 8.4 percent of households that have three or more adults in AFDC-UP (two-parent) cases or have two or more adults in AFDC-FG (one-parent) cases at the time of sampling because we are unable to construct useful decision rules about which adult can claim a child for purposes of the EITC. We also exclude a small number of cases where we were unable to determine the date at which they entered the sample, where adults were missing a birth date, where the gender or race/ethnicity of the adults were missing, and where no children were present in the case. Overall, we are left with 626,700 cases, or 82.5 percent of an original sample of 759,913 cases.

Our analysis sample defines a single "sampling date" – the 4th quarter in a random year of the household's welfare spell – for each of the 626,700 useable cases in the Rand sample.[19] We

[17] The original Rand sample was intended to be a 17 percent sample of adults and children on welfare. But once a person was drawn into the Rand sample, all people in all of the sampled person's cases were included. The resulting sample reflects roughly (depending on the year) 50 percent of all adults ever on welfare.

[18]We define a child as being 18 years old or younger throughout the year. This age limit corresponds to the age of a qualifying child for EITC purposes. These child-only cases occur when the children are eligible for AFDC but the adults who live with the children are ineligible.

[19] Note that each case in the Rand sample only enters our analysis sample once, using information on that case as of the (randomly) selected sampling date drawn from each case's time on welfare during the 1990s.

do this for two reasons: first, we only have demographic information, particularly information on the number and ages of children in the household, for periods that the household is receiving welfare. The information is critical to us, since our identification strategy is predicated on examining employment differences between families with one and with two or more children. By focusing on a single sampling date, we treat households that stay on welfare (for whom we have ongoing demographic information) symmetrically to households who do not stay on welfare (for whom we do not have ongoing demographic information). Second, we also do not want our analysis to be overly influenced by the behavior of long-term welfare recipients, as would happen if we sampled with replacement from the 626,700 useable cases in the Rand sample.

When creating our analysis sample, we have to determine how many years of employment data we want to use around the sampling date. Our central analyses use two variants. The first is a conceptually straightforward cross-sectional sample, where we examine employment in the year following the sampling date. Because we use only one year of employment data, analyses with the cross-sectional sample do not include household fixed effects.

Our longitudinal sample balances three considerations. First, the longer the time span before and after the sampling date, the greater the chance we will make errors in assessing the number of children in the family. The errors are likely to be asymmetric when going backward in time relative to going forward. Specifically, we know the ages of children at the time of sampling so we can precisely determine their date of birth. We do not observe births occurring after the sampling date. Therefore, information on the ages and number of children in the household will likely be more accurate using data prior to the sampling date compared to data after the sampling date.

Second, more data, all else being equal, increase the precision of the longitudinal estimates.

Third, with fixed effect models of employment or EITC claiming, households that change status from one child to two or more children or *vice versa* identify the coefficients of interest. This status changes for two reasons in our sample: children are born, or their age exceeds the EITC qualifying child age limits. We observe births in the data as we examine employment in years prior to the sampling date. We observe aging out as we examine employment in years following the sampling date.

In extensive descriptive analyses available from the authors on request, but not included in the paper, it is clear that households with children that age out of the EITC qualifying child age range have substantially different *observable* characteristics than the typical (or mean) household on welfare in California: they are older, have more children, have different racial and ethnic composition and have different earnings. Given these differences in observable characteristics, we think it is likely that there are substantial unobserved differences in households with one and two or more children when estimates depend importantly on households with children that age out of the EITC qualifying child eligibility ranges. In contrast, families with children born in the four years prior to the sampling date are indistinguishable (besides having a young child) from all families on welfare.

Our primary longitudinal sample balances the three considerations mentioned above. It covers a 4-year period from 3 years prior to the date of sampling through the year in which the household is sampled. We focus on AFDC-FG (single parent) cases, which account for 84.1 percent of the total cases (or 527,125). We conduct sensitivity analyses using the AFDC-U (two-parent) cases.

Appendix Tables 1a and 1b shows the weighted characteristics of the primary cross-sectional (Table 1a) and longitudinal analysis samples (Table 1b). Overall average employment

rates were around 48 percent. In detail not shown in the table, employment rates in the cross-sectional sample increase from 36.3 percent in 1991 to 65.4 percent in 2000. Recall that *every* adult in these cross-sectional cases was on welfare in the year prior to when we record employment, so the economy, the EITC, welfare reform, or other factors greatly increased employment of welfare recipients over the period examined. The racial and ethnic characteristics of the population changed over time as well – the proportion of cases headed by a Hispanic increased 15 percentage points and the proportion of cases headed by a Black increased 3 percentage points, while the proportion of White families fell by roughly 18 percent points. Other features of the caseload remained roughly constant over time. More than half the cases in the sample have two or more children, while roughly 30 percent have exactly two children.

3. Employment Rates and EITC Claiming by Family Size and Year in the Cross-Sectional Analysis Sample

In this section, we present unadjusted (for covariates) difference-in-differences estimates of household (adult) employment and EITC claiming. Let Y_{ict} denote an outcome variable for the adults in the i^{th} household from county c in (calendar) year t. In our case, there are two outcome variables, whether one or more adults in a household are employed in the labor market (Emp_{ict}) and whether a household files a tax return and claims the EITC ($ClaimEITC_{itc}$), where these two outcomes are defined as follows:

$$Emp_{ict} = \begin{cases} 1, \text{ if at least 1 adult in } i^{th} \text{ household in county } c \text{ is employed in year } t \\ 0, \text{ otherwise} \end{cases}, \quad (1)$$

$$ClaimEITC_{ict} = \begin{cases} 1, \text{ if } i^{th} \text{ household in county } c \text{ files tax return \& claims EITC in year } t \\ 0, \text{ otherwise} \end{cases} \quad (2)$$

As noted in the Introduction, we use the differential change in the EITC by family size before and after 1994 to identify the EITC's employment effect and the EITC's effect on tax

filing *and* EITC claiming. In particular, we use the following "difference-in-difference" estimator:

$$\beta_s^* \equiv \left(\overline{Y}_{t''}^{2+} - \overline{Y}_{t''}^{1}\right) - \left(\overline{Y}_{t'}^{2+} - \overline{Y}_{t'}^{1}\right), \tag{3}$$

for $t'' \geq 1994 > t'$, where $\overline{Y}_s^1$ is the mean outcome in year *s* for households that had only one child and $\overline{Y}_s^{2+}$ is the corresponding mean for households that had two or more children.[20] Absent cohort, welfare entry, or other sample compositional effects – a major qualification that we address in the following sections – we expect, if the EITC stimulates employment and/or rates of filing taxes and claiming the EITC, to see higher rates of these outcomes for two-plus child families relative to one-child families as the EITC differential between the two types of families begins to increase substantially in 1994.[21]

We first present estimates for employment rates in Table 2a. The "Diff-in-Diff" column of Table 2a suggests that the EITC expansion beginning in 1994 may have had a significant, positive effect on the employment rates, though if the EITC is the *causal* mechanism, either its influence is occurring with a considerable lag or other factors not accounted for in the table mask the EITC's effect. In particular, employment rates for both one child families and for families with two or more children are steady in the 1991-93 base period and then begin to rise sharply in 1994. However, the employment increases for families with one child and for families with two or more children are equivalent until 1997, the year after the EITC changes are fully phased in. The differences in employment rates for families with two or more children and one-child

[20] Each year we include a household in the sample if an adult (defined as being 18 or older) and at least one child (defined as being under 18) are present. Based on birth dates, we compute ages of household members each analysis year and reclassify a household appropriately. For example, consider a woman with two children born in 1992 and 1994. Prior to 1992, we drop this household since no children existed. We classify the household as having a single child from 1992 to 1994 and as having two children after that. If the panel were longer, in 2011 the eldest child would be too old and the household would revert to a one-child classification.

[21] The full $1,404 maximum differential in the EITC by number of children does not occur until 1996.

families are large and highly statistically significant beginning in 1998. Of course, these patterns may be the result of other factors changing over this period, including the economy, the welfare system, and compositional issues related to entry and exit of our sample.

A unique feature of our analysis is having access to the individual tax returns household file. Table 2b shows a similar set of difference-in-difference calculations where the outcome is filing a tax return and claiming the EITC. The temporal pattern of EITC claiming is similar to the pattern for employment. Households with two or more children are between 2 and 2.5 percentage points *less* likely to claim the EITC from 1991 through 1995. This differential starts narrowing in 1996 and by 2000 families with two or more children are 4.2 percentage points *more* likely to claim the EITC. As with the employment estimates, many things could account for the patterns of EITC filing that we observe.

We refine our analysis of employment and claiming the EITC in the next section.

4. Multivariate Analysis of Employment and EITC Claiming

The unadjusted difference-in-difference estimates presented in Tables 2a and 2b may not isolate the effects of the EITC on employment or EITC claiming for at least three related reasons. The first concerns the fact that the number of children a household has may have a separate influence on these outcomes over and above the influence of the EITC. Second, and related to the first, economic models of household time allocation and fertility decisions imply that the number of children (and their age-composition) and employment outcomes are jointly determined by observable and possibly unobservable factors that determine household's tastes, home and market productivity, etc.[22] Third, there are potentially other policy changes, most importantly changes in welfare policy, that occurred around the time of the EITC expansion in

[22] See, for example, Hotz and Miller (1988).

the 1990s that may have affected the employment decisions of households with differing number of children. In an attempt to account for factors that may compromise the validity of the unadjusted EITC effect estimates in Tables 2a and 2b, we also present several alternative sets of "adjusted" difference-in-difference estimates using regression methods.

The first set of adjustments is based on the following regression function:

$$Y_{ict} = \phi + \sum_{s=1992}^{2000} \alpha_S Year\_s_t + \sum_{s=1994}^{2000} \beta_s [2{+}Kids_{ict} \cdot Year\_s_t] + \sum_{j=1}^{9} \delta_j Kids\_j_{ict} + \sum_{k=1}^{17} \psi_k KidsAge\_k_{itc} + \kappa X_{ic} + \lambda W_{ct} + \eta L_{ct} + \sum_{m=1}^{55} \theta_m County\_m_c + \varepsilon_{ict} \quad (4)$$

where *Year_s* is an indicator variable equal to 1 if the current year *t* equals *s* and 0 otherwise, *s* = 1992,…,2000; 2+$Kids_{ict}$ is an indicator variable equal to 1 if the household has two or more children in year *t*; $Kids\_j_{ict}$ is an indicator variable equal to 1 if there *j* children in the household in year *t* and 0 otherwise, *j* = 1,…,9 where the category '9' is for 9 or more children; $KidsAge\_k_{ict}$ = the number of children in the household that are age *k* in year *t*, *k* = 1,…,17; X_{ic} denotes a vector of time-invariant demographic characteristics of the i^{th} household; W_{ct} denotes a vector of county-level measures of California's welfare caseload and policies affecting this caseload; L_{ct} denotes a vector of time-varying indicators of county-level labor market conditions; $County\_m_{ict}$ is an indicator variable equal to 1 if the i^{th} household resides in county *m*, *m* = 1,…,55; and ε_{itc} is an independently and identically distributed error term.

The parameters, β_s, in (4) denote year-specific differential effects of having two or more children on the outcome and, thus, are the adjusted (for covariates) version of the unadjusted difference-in-differences estimator, (β_s^*), defined in (3) and for which estimates are presented in Tables 2a and 2b. Estimates of these and the other parameters in (4) can be obtained by estimating (4) by ordinary least squares (OLS), using the cross-sectional data described in

Section 2.2.[23]

Using the longitudinal data, we also estimate a generalization of the regression specification in (4), modifying it to allow for household-specific fixed effects as follows:

$$Y_{ict} = \phi_i + \sum_{s=1992}^{2000} \tilde{\alpha}_S Year\_s_t + \sum_{s=1994}^{2000} \tilde{\beta}_s [2{+}Kids_{ict} \cdot Year\_s_t] + \sum_{j=1}^{9} \tilde{\delta}_j Kid\_j_{ict} + \sum_{k=1}^{17} \tilde{\psi}_k KidsAge\_k_{itc} + \tilde{\lambda} W_{ct} + \tilde{\eta} L_{ct} + \sum_{m=1}^{55} \tilde{\theta}_m County\_m_c + \varepsilon^*_{ict}. \quad (5)$$

The time-invariant households variables (X_{ic}) are not included in (5) as these factors will be absorbed by the ϕ_is. In essence, controlling for the ϕ_is allow us to account for *all* time-invariant characteristics of the household, such as the educational attainment, work readiness, motivation, etc., that might affect household employment decisions and the likelihood of filing a tax return and claiming the EITC.

We note that both of the outcomes we analyze, *Emp* and *ClaimEITC*, are dichotomous dependent variables so that the specifications in (4) and (5) characterize linear probability models. [24] In both of these specifications, while we assume that ε_{ict} is independent across households, we do adjust the standard errors on the regression coefficients for correlations within a household across time.

Two crucial assumptions are needed for our empirical model to be sensible. First, any independent effects of being in the "treatment" or "control" groups on employment do not vary temporally. Put differently, one-child and multi-child families (the building blocks of the treatment and control groups) respond identically to changes in economic conditions and welfare

[23] Note that one also estimate (4) using the longitudinal data on households described in Section 2.2.

[24] We present fixed effect linear probability results in our longitudinal analyses since they are particularly easy to interpret and they are computationally less demanding given our large samples than panel logits or alternative approaches.

policies over the period being analyzed after conditioning on other characteristics.[25] We include an extensive set of covariates to increase the likelihood that this condition holds, and examine the robustness of our results across alternative specifications. We also implement stringent specification checks that are described below. Second, we assume that the distribution of employment propensities (conditional on covariates) for the populations of one-child and multi-child families remain stable before and after the EITC expansion. Our longitudinal analysis mitigates, if not eliminates, this potential problem, at least relative to papers in the existing literature.[26]

We now briefly describe the actual variables that compose the vectors X_{ic}, W_{ct} and L_{ct} included in the regression specifications in (4) and (5).

4.1 Other Influences of Children and Household Characteristics

Employment rates tend to increase as people and children age. We include a detailed set of covariates to account for the composition and characteristics of the family. We include 18 variables counting the number of children at each age, from 0 to 17, in the household in the particular analysis year. We also include a set of dummy variables corresponding to the number of children in the household in each year. In empirical specifications that exclude household fixed effects, we also include indicator variables for female, Hispanic, Black, Asian, or mixed or other races or ethnicities (leaving whites as the excluded category). We also include a variable for the mean age of adults in the household.

[25] See, Heckman and Robb (1985) and Moffitt and Wilhelm (2002) for further details on difference-in-difference estimators.

[26] With repeated cross-section data, maintaining the second assumption is more problematic. It requires the distribution of employment propensities (conditional on covariates) for the populations of childless, one-child and multi-child single women to remain stable over the period being analyzed. To the extent that different groups of households enter the sample before and after the intervention and these groups differ, even if only with respect to their time-invariant employment propensities, the assumption is unlikely to hold with repeated cross-section data.

4.2 Accounting for the Potential Influences of California's Welfare Programs

Beginning in October 1992 California, under a series of waivers granted by the U.S. Department of Health and Human Services, altered its AFDC program.[27] Judged by the standards of welfare experimentation and reform elsewhere, the changes were fairly modest. The most important change was to reduce the maximum amount of AFDC cash aid by 15 percent.[28] Static models of labor supply predict, under usual assumptions about income and substitution effects, that the combination of the reduction in the guaranteed levels of AFDC benefits and other features of the package will tend to increase employment. Evaluations of California's welfare changes, however, suggest the effects of the changes (collectively known as "California Work Pays") were modest (Becerra, et al. 1998; Hotz, Scholz, and Mullin, 2002a, 2002b).[29] Consequently, welfare changes seem an unlikely candidate for explaining substantial changes in employment among California's welfare population.

Nevertheless, we include in the empirical model a measure summarizing county-level enrollments as a percentage of each county's AFDC caseload in California's welfare-to-work program, Greater Avenues for Independence (GAIN). Counties could implement different GAIN

[27] Hotz, Mullen, and Scholz (2002a, 2002b) evaluate the effects of these changes on employment, income, and welfare use.

[28] The other changes were also designed to encourage employment. These changes (a) eliminated the 100 hour (per month) work limitation once a household started receiving AFDC-U; (b) eliminated the time limit for the $30 and 1/3 income disregard; (c) encouraged pregnant teens and teen parents to stay in or return to school by providing child care, transportation, and other assistance, and created disincentives for bad grades or for dropping out of school; (d) raised (modestly) asset limits for families once they start receiving benefits; (e) allowed AFDC-eligible persons to decline an AFDC cash grant, but still receive Medi-Cal (the Medicaid program in California) and child-care assistance; and (f) made California's Job Opportunities and Basic Skills (JOBS) training program and the Greater Avenues for Independence (GAIN) program more work-oriented.

[29] Meyers et al. (1998) examine implementation of California Work Pays based on 66 intake or recertification interviews in 1993 and 1994. Their evidence suggests little was done to convey the broad message of work pays to clients. They write, "Only one office had a Work Pays poster within sight of clients. Two workers were observed to have Work Pays buttons. When asked about the button, however, one of the two reported that she had obtained it from her union – as part of a campaign to make 'work pay' for county employees" (p. 12).

programs and were free to change them over this period.[30]

In the cross sectional analysis we also include a set of indicator variables for each entry cohort onto welfare. These control for differences in labor force participation associated with the passage of time after initial AFDC receipt. They also help account for differences in entry cohorts across years that may be related to economic conditions. These indicator variables drop out of the longitudinal analyses, since the household-specific fixed effect accounts for time invariant characteristics of households.

Low-income households in California (and elsewhere) are potentially eligible for other social assistance programs that may affect their decisions to work. These other programs, such as Medi-Cal, Food Stamps, and SSI, did not change much or at all over this period. They also generally did not vary across California's 58 counties. Thus, we do not attempt to examine the effects of other programs on labor force participation, beyond including year dummies in the analysis.[31]

In sensitivity analysis we add a variable showing county-level spending by year in state-provided child care assistance. Public child care expenditures were small in California in the early 1990s, but increased sharply late in the decade (Marrufo, O'Brien-Strain, and Oliver, 2003). These expenditures could influence employment patterns, so we want to be sure that our results are not a result of changes in child care spending.

4.3 Accounting for the Influence of Local Labor Market Conditions

Some of the changes in welfare caseloads and employment rates of low-skilled workers

[30] For more on the GAIN program and the differences in its implementation in California's counties, see Riccio, et al. (1989) and Hotz, Imbens, and Klerman (2002).

[31] In the empirical models using cross sectional data we also include a full set of county dummies. The results are also unaffected by including a full set of county-by-year interaction variables, which account for county-specific trends.

during the 1990s may be attributable to changes in labor market conditions. To account for the role of the economy, we examine several labor market characteristics, including the overall employment rate in the 58 California counties; the share of employment in manufacturing, service, and retail trade (the employment share in government services is the excluded category); and income per worker in each county, and income per worker in each sector in the county (again, excluding government services).

There is substantial cross-county variation in local economic conditions across California. We summarize these trends and differences by graphing the annual employment to population ratios across California regions in Figure 1.[32] Employment-to-population ratios are less than 55 percent in the Northern and Mountain counties of California (they are less than 35 percent in Calaveras County), while they exceed 75 percent by the late 1990s in the Bay Area. Statewide, average county-level employment-to-population ratios fell from 62 percent to 61 percent during the recession early in the decade and increased to 68 percent by 2000.

Figure 2 presents similar statistics, showing annual real earnings (in $1,000s of 1992/93 dollars) per service sector worker across California regions. Average annual earnings in the mountain and northern counties ranged from $20,700 to $25,000, while they range between $37,100 and $52,300 in the Bay Area. Statewide average annual earnings range between $33,200 and $38,900. These figures document the unusually large variation in economic conditions across California counties.

As mentioned above, we also include indicator variables for years in our empirical model to account for aggregate business cycle effects.

As mentioned previously Appendix Tables 1a and 1b provides a summary list and

[32] Similar figures for disaggregated sectors are available on request. These data were obtained from the website of the Labor Market Information Division (LMID) of the State of California's Economic Development Department.

descriptive statistics for the variables that we use in the *X*, *L*, and *W* vectors in the estimation of the regression models in (4) and (5).

5. Identifying the Effects of the EITC

Our first test examines employment differences between families with two or more children relative to one-child families, using fixed effect empirical models with additional time-varying household and county-level covariates. If the EITC increases employment, we expect to see employment for families with two or more children rise relative to the employment of one-child families after 1993.

If the EITC *causes* differential employment increases, families should file tax returns and claim the EITC. Our second test examines this intuition by estimating fixed effect empirical models of households claiming the EITC.

Our third test examines whether employment changes apply to all families with two or more children relative to one-child families, as would be expected if EITC changes increase employment. If other factors cause differential employment increases across families of different sizes, we would expect differences to also arise from families with two children relative to families with three or more children. We examine this intuition by focusing on the specification check where we examine how employment for families with exactly two or more children differs from the employment of families with three or more children.[33]

If the EITC *causes* differential employment increases, we would expect them to be present *only* for families who file tax returns. Our fourth test examines this intuition by conditioning the sample of households who *do not* file returns: we do not expect to see employment increase over

[33] Recall, the test of whether there is a difference between the effects for families with two or more children relative to the effects for families with three or more children is whether the coefficient on the specification check term is zero for each year and whether they are jointly equal to zero across years.

time for families with two or more children relative to one-child families in this alternative sample.

We examine the results of these four tests as well as sensitivity analysis below.

6. Empirical Results

6.1 Estimates of EITC Effects for Multivariate Models

Table 3 presents results for our central OLS and fixed effects regression specifications (all specifications have a similar structure, though specifications without household fixed effects include additional covariates, as noted above). The dependent variable indicates whether or not the adult in the household participated in the labor market (i.e., had positive EDD earnings) during the year. In the OLS specification (estimated with the cross sectional sample), employment is measured in the year following the date of sampling (which occurs in the 4^{th} quarter of the previous year). Each OLS regression includes an extensive set of covariates shown in Appendix Table 1a. The disadvantage of the OLS specification is that it does not include the household-specific effect. The advantage is that there are likely to be fewer errors in measuring the number and ages of children in the household and other household characteristics. We prefer the fixed effect estimates, but finding results consistent with the longitudinal analysis with the cross-sectional data would strengthen our interpretation.

The results for the first test mirror the unconditional employment estimates given in Table 2a. Reading on the right-hand side of the table (the fixed effect estimates), the estimates on the variables "2+ Kids in *Yr*" for *Yr* = 1994,…,2000, show that employment rates for those with two or more children are statistically indistinguishable from the employment rates of one-child families in 1994 and 1995, conditioning on the fixed effect and other covariates. Employment rates for families with two or more children are 1.2 percentage points higher in 1996, 2.6

percentage points higher in 1997, 3.2 percentage points higher in 1998, 3.4 percentage points higher in 1999, and 3.0 percentage points higher in 2000, relative to employment rates for one-child families.

The OLS results in Table 3 are similar to the previous results, though they provide somewhat weaker evidence of positive EITC employment effects. Employment for families with two or more children increases significantly in 1998 through 2000 relative to employment of families with one child.

Results for the second test are given in Table 4, which shows OLS and fixed effect regression estimates of EITC claiming. The estimates from both models mirror the employment results from Table 3. EITC claiming by families with two or more children is insignificantly different from EITC claiming by one-child families in 1994 and 1995. Between 1996 and 2000 families with two or more children are significantly more likely to claim the EITC than are families with one child (the fixed effect estimates are only marginally significant in 2000 and the OLS estimates are insignificant in 1997). If the EITC is responsible for the employment differentials between families with two or more children compared to one-child families, the Table 4 estimates reflect the pattern of EITC claiming we would expect to see.

Other correlates of employment and EITC claiming. The empirical models in Tables 3 and 4 include a number of additional correlates of employment and EITC claiming (complete results are available on request). There are very strong year effects, with employment increasing by 24 percentage points from its low point in 1993 to its high point in 2000. EITC claiming increases by 35 percentage points between 1992 and 2000. As shown in Table 1, the EITC increased substantially over the 1990s both in levels (for all families with children), as well as increasing disproportionately for larger families (which is the focus of this paper). Along with the broad

EITC expansion, groups around the country engaged in considerable outreach.[34] The fact that EITC claiming increased more rapidly than employment suggests that the combination of larger potential EITC payments along with outreach efforts induced more people to claim the credit.

The family size variables have opposite effects on employment and EITC claiming. Employment increases steadily with the age of the children. Relative to the excluded category of 18-year-old children, households with a child under 2 are 9 to 16 percentage points less likely to work. GAIN enrollments appear to increase employment and EITC claiming, but the economic magnitude of the estimated correlation is small. Conditioning on other covariates, counties with 100 percent of their caseload enrolled in GAIN have employment rates (EITC claiming rates) 1.6 (2.3) percentage points higher than counties where nobody enrolled in GAIN.

Employment and EITC claiming are affected symmetrically by local labor market conditions. Both, not surprisingly, are positively correlated with county employment-to-population ratios and with the average earnings per worker in the county of residence. Residents of counties with high average manufacturing sector earnings have low employment and EITC claiming rates, suggesting that manufacturing companies are not primary employers of households in our sample.

6.2 Tests Results and Other Sensitivity Analyses[35]

The third test focuses on whether the differential employment trends observed in the data for one and two-or-more child families also holds for families with three or more children compared to two-child families. If the patterns also hold for larger families, then some factor other than the

[34] The Center for Budget and Policy Priorities, for example, has had a long-standing national EITC outreach campaign. For a recent example, see http://www.cbpp.org/eic2005/. There are many other examples from every part of the country.

[35] Complete results for every specification discussed in this section are available from the authors on request. In the interest of space, we provide only summary results in the tables and discussion that follows.

EITC likely is causing the employment differences documented in Table 3.

Column (1) in Table 5A repeats the fixed effect estimates from Table 3. Column (2) shows the crucial "specification check" estimates. The coefficients on the [2+ Kids – 3+ Kids in *Yr*] variables in column (2) are individually insignificant. Moreover, the bottom of the table reports the F-test, showing they are jointly insignificant. Thus, if something other than the EITC causing employment rates of families with two or more children to increase by 3 percentage points relative to employment rates of one-child families in the second half of the 1990s, the alternative factor must not differentially affect families with three or more children relative to two-child families. It is hard to think of what this alternative would be.

In the fourth test, we restrict the sample to households that do not file tax returns. The sample is sizeable, with 564,612 observations.[36] Moreover, while these households do not file tax returns, a substantial fraction are employed on average in any given year (in the overall sample around 45 percent are employed). The results for this test are given in column (3) of Table 5A. There are no employment differences between families with one child, two children, or three or more children. Thus, whatever is causing increases in the employment of families with two or more children relative to one-child families, it only occurs for those who file tax returns.

Table 5B shows similar patterns for EITC claiming.

We think the most sensible interpretation of the evidence from the four-part test is that the EITC is responsible for most of the differential employment increase of families with two or more children relative to one-child families.

[36] Of course in the fixed effects specification, a given household can appear in the sample as many as four times. All standard errors reported in the paper are adjusted for the fact that the error term will not be independently distributed for observations that arise from the same household.

6.2.1 Altering the base year definition

In all specifications presented to this point we use average employment or average EITC claiming between 1991 and 1993 as the reference period for the regressions. We do this because the maximum EITC differential between families with two or more children and one child was inconsequentially different from $0 during these years. But pooling the reference period masks another subtle specification check for our model. Namely, there should be no significant employment (or EITC claiming) trends prior to the period when the EITC began to diverge for the two groups of households. Unfortunately, because data are not available prior to 1991, we cannot go back further to examine employment and EITC claiming trends.

There is absolutely no evidence of employment trends between 1991-1993. In fixed effects employment specifications shown in Table 5C, for example, the coefficient (standard error) on the "2+ Kids in 1991" variable is -.004 (.004) and the coefficient (standard error) on the [2+ Kids – 3+ Kids in 1993] variable is .001 (.004).

6.2.2 Altering the definition of employment

Readers may not think earning, say, $10 a year in the labor market (where earnings are reported to the UI system) is a very consequential achievement when thinking about the economic importance of the employment patterns documented in this paper. This concern arises because we define employment as having *any* earnings reported to the UI system in a given year. To address this concern, we estimated our fixed effect empirical models defining employment as having earnings exceeding $500 in the year (the left-hand panel of Table 5D), and as having earnings exceeding $1,000 in the year (the right-hand panel of Table 5D).

The patterns are similar to the baseline estimates. The consequence of increasing the earnings threshold above which we define a household as being "employed" is the magnitude

and the significance of the coefficient estimates decreases somewhat as the threshold is increased.

6.2.3 Focus on AFDC-UP cases

It is well known that the EITC can have negative employment incentives on the intensive margin for married couples if one partner's earnings place the couple in the phaseout range of the credit (see, for example, Eissa and Hoynes, 2004). For our analysis, however, we define employment as the first dollar of the *couples* earnings, so by this definition, the credit still has unambiguously positive employment incentives (on the extensive margin). Roughly 70 percent of the AFDC-UP cases have some earnings in a given year.

We find no evidence that the differential expansion of the EITC for families with two or more children increased employment of adults in AFDC-UP households. In the fixed effects empirical models including the specification check terms (the "2+ Kids" by year interaction terms), we find a negative, significant coefficient of –0.056 on the "2+ Kids in 1999" variable and positive and significant terms on "2+ Kids in 1999" and "2+ Kids in 2000". This indicates that the economy or welfare reform or some other factor began to affect the employment of AFDC-UP families late in the 1990s, and these influences had differential effects for families of different sizes. But the patterns are not what one would expect to see if the EITC were the factor causing the changes.

6.2.4 Can changes in childcare spending explain the employment trends for single-parent families?

Public expenditures for childcare were quite low for most of the period we examine in the paper. But late in the 1990s, state childcare spending increased sharply (Marrufo, O'Brien-Strain, Oliver, 2003). This is also the period when the employment of single parents with two or more children increased employment relative to parents with one child. Moreover, the increases come

after the differential EITC expansions, which raises the question of whether some other factor is causing the patterns observed in the data. To some extent we have already discussed this issue, because the changes would need to affect employment of families with two or more children relative to one-child families, but not affect employment of families with three or more children relative to two-child families. It is unlikely that childcare expenditure increases would have this empirical pattern. But it is nevertheless straightforward to include a measure of county (public) childcare expenditures for the welfare caseload in our empirical fixed effects regression models of employment.

Two results stand out from this alternative specification. First, the central coefficients on employment are unchanged (they actually get slightly larger). The coefficients on the specification check terms remain insignificant, individually and jointly. Second, the estimated coefficient on the childcare variable is consistently *negative* and precisely estimated. We think the most likely explanation for the result is a "policy endogeneity" phenomenon. Counties with more severe employment problems, particularly in the late 1990s after a decade-long economic expansion, implemented more aggressive childcare subsidy programs. Thus we find a negative correlation between childcare spending and employment. An expansion of childcare spending, however, does not account for the employment patterns documented in this paper.

7. Conclusions

The EITC transfers a large amount of money to working poor families and reduces poverty. There is also a considerable amount of evidence that the credit not only redistributes resources, but also encourages employment, thereby avoiding one of the negative behavioral incentives of traditional income transfer programs. No previous EITC employment study, however, has used data on whether or not people eligible for the EITC actually file tax returns and claim it. Tax data

reinforce the finding that the EITC appears to positively influence employment. Patterns of EITC-claiming appear to mirror employment patterns in the data. The effects appear only when comparing families with two or more children relative to one-child families, rather than families with three or more children relative to two-child families. This pattern rules out alternative explanations that would have differential employment effects as families have more children. Moreover, there are no EITC-like employment effects in samples where households do not file tax returns.

In addition to having access to tax data, our work also adds several distinctive features to the EITC employment literature. We use a different identification than most papers on the topic, where we compare employment rates of families with two or more children to employment rates of one-child families over a period where the EITC increases more rapidly for families with two or more children. We estimate our empirical model with longitudinal data, which allow us to account for time-invariant characteristics of households that may affect employment. And our analysis is based on California, which allows us to account in a detailed way for local labor market and transfer program characteristics that likely affect employment. Lastly, we focus on the portion of the population that is of considerable interest to policymakers: women with children who are trying to make the transition from welfare to work.

The estimates from Table 3 suggest that the differential EITC expansions increased employment by as much as 3.7 percentage points for families with two or more children relative to families with one child. As shown in Table 2, overall employment rates increased by 26.6 percentage points for one-child families and by 31 percentage points for families with two or more children. The average employment increase for the two groups was 28.8 percentage points: the differential EITC accounted for 12.8 percent of the average increase in employment over this

period. This would seem somewhat smaller than what is implied by the Meyer-Rosenbaum (2001) and Grogger (2003) results.[37] But this illustrative calculation may provide an unduly pessimistic view of the EITC. The differential EITC expansion can explain, in the regression-adjusted fixed effect specification, 84 percent of the *difference* in the change of employment between families with two or more children and one-child families.

A different approach to assessing the magnitude of the empirical estimates may provide a better perspective of the EITC's effects. Hotz, Scholz, and Mullin (2002a) report that average annual disposable income for single-parent welfare cases in California in 1998 was around $10,000. The average EITC differential for families with one and two or more children in the sample was roughly $480 in 1998. Hence, the EITC increased disposable income by roughly 4.8 percent. Employment rates in 1998 were around 61 percent, and the EITC expansion increased employment of families with two or more children by 3.3 percentage points, or 5.4 percent. This implies an employment elasticity with respect to disposable income of 1.1, which is at the upper end of the range of estimates discussed in Hotz and Scholz (2003).

This work suggests that the EITC can be an important tool in efforts to increase employment of welfare recipients.

[37] As mentioned earlier, Grogger (2003) concludes that the EITC "may be the single most important policy measure for explaining the decrease in welfare and the rise in work and earnings among female-headed families in recent years" (p. 408). Meyer and Rosenbaum (2001) find that EITC changes account for roughly 60 percent of the increase in the employment rate of single mothers from 1984 to 1996 and roughly 31 percent of the increase from 1992 to 1996.

References

Bartik, Timothy J. and Randall W. Eberts, 1999, "Examining the Effects of Industry Trends and Structure on Welfare Caseloads," W.E. Upjohn Institute for Employment Research, February.

Becerra, Rosina M, Vivian Lew, Michael N. Mitchell and Hiromi Ono, 1998, *Final Report: California Work Pays Demonstration Project, Report of the First Forty-Two Months*, School of Public Policy and Social Research, University of California, Los Angeles, October.

Cancian, Maria, and Arik Levinson, 2003, "Labor Supply and Participation Effect of the Earned Income Tax Credit: Evidence from the National Survey of America's Families and Wisconsin's Supplemental Benefit for Families with Three Children," mimeo, University of Wisconsin -- Madison and Georgetown University, http://www.georgetown.edu/faculty/aml6/pdfs&zips/Cancian-Levinson.pdf.

Dickert, Stacy, Scott Houser and John Karl Scholz, 1995, "The Earned Income Tax Credit and Transfer Programs: A Study of Labor Market and Program Participation," *Tax Policy and the Economy*, James M. Poterba (ed.), National Bureau of Economic Research and the MIT Press, 9, 1-50.

Eissa, Nada and Hilary Williamson Hoynes, 2004, "Taxes and the Labor Market Participation of Married Couples: the Earned Income Tax Credit," *Journal of Public Economics*, 88, 1931-1958.

Eissa, Nada and Jeffrey B. Liebman, 1996, "Labor Supply Response to the Earned Income Tax Credit, *Quarterly Journal of Economics*, May, 605-637.

Ellwood, David T., 2000, "The Impact of the Earned Income Tax Credit and Social Policy Reforms on Work, Marriage, and Living Arrangements," *National Tax Journal*, Volume LIII, No. 4, Part 2, December, 1063-1105.

Grogger, Jeffrey, 2003, "The Effects of Time Limits, the EITC, and Other Policy Changes on Welfare Use, Work, and Income Among Female-Headed Families," *Review of Economics and Statistics*, May, 85(2), 394-408.

Heckman, James J., "Comment," in *Empirical Foundations of Household Taxation*, Feldstein and Poterba (ed.), NBER and University of Chicago Press, 1996, 32-38.

Heckman, James J. and R. Robb, 1985, "Alternative Methods for Evaluating the Impact of Interventions," in J.J. Heckman and Burton Singer (eds.), *Longitudinal Analysis in Labor Market Data*, Econometric Society Monographs Series, Cambridge University Press.

Hotz, V. Joseph and Robert A. Miller, "An Empirical Model of Life Cycle Fertility and Female Labor Supply," *Econometrica*, 56(1), 91-118.

Hotz, V. Joseph, Charles H. Mullin, and John Karl Scholz, 2002a, "The Effects of Welfare Reform on Employment and Income: Evidence from California," mimeo, http://www.ssc.wisc.edu/~scholz/Research/CWPDP_paper_version6.pdf.

Hotz, V. Joseph, Charles H. Mullin, and John Karl Scholz, 2002b, "Welfare, Employment and Income: Evidence on the Effects of Benefit Reductions in California," *American Economic Review Papers and Proceedings*, May, 2002, 380-384.

Hotz, V. Joseph and John Karl Scholz, 2002, "Measuring Employment and Income Outcomes for Low-Income Populations with Administrative and Survey Data," in *Studies of Welfare Populations: Data Collection and Research Issues*, National Research Council, National Academy Press, 275-315.

Hotz, V. Joseph and John Karl Scholz, 2003, "The Earned Income Tax Credit," in *Means-Tested Transfer Programs in the United States*, Robert Moffitt, Editor, The University of Chicago Press and NBER, 141-197.

Hotz, V. Joseph, Guido Imbens, and Jacob Klerman, 2002, "The Long-Term Gains from GAIN: A Re-Analysis of the Impacts of the California GAIN Program," Unpublished manuscript, UCLA, http://www.econ.ucla.edu/hotz/working_papers/gain.pdf.

Keane, Michael and Robert Moffitt, 1998, "A Structural Model of Multiple Welfare Program Participation and Labor Supply," *International Economic Review*, 39(3), 553-589.

Marrufo, Grecia, Margaret O'Brien-Strain, and Helen Oliver, 2003, "Child Care Price Dynamics in California," mimeo, Public Policy Institute of California, http://www.ppic.org/content/pubs/R_1203GMR.pdf.

Meyer, Bruce D. and Dan T. Rosenbaum, 2000, "Making Single Mothers Work: Recent Tax and Welfare Policy and Its Effects," *National Tax Journal*, Volume LIII, No. 4, Part 2, December, 1027-1061.

Meyer, Bruce D. and Dan T. Rosenbaum, 2001, "Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers," *Quarterly Journal of Economics*, CXVI(3), August, 1063-1114.

Meyers, Marcia K., Bonnie Glaser and Karin MacDonald, 1998, "On the Front Lines of Welfare Delivery: Are Workers Implementing Policy Reforms?" *Journal of Policy Analysis and Management*, 17(1), 1-22.

Moffitt, Robert A., and Mark O. Wilhelm, 2000, "Taxation and the Labor Supply Decisions of the Affluent," *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*, eds. Slemrod, Joel, Russell Sage Foundation & Harvard Press, New York.

Riccio, J., B. Goldman, G. Hamilton, K. Martinson, and A. Orenstein (1989), *GAIN: Early Implementation Experiences and Lessons*, Manpower Demonstration Research Corporation, April.

Figure 1: Annual Employment to Population Ratios, California, 1992-2000

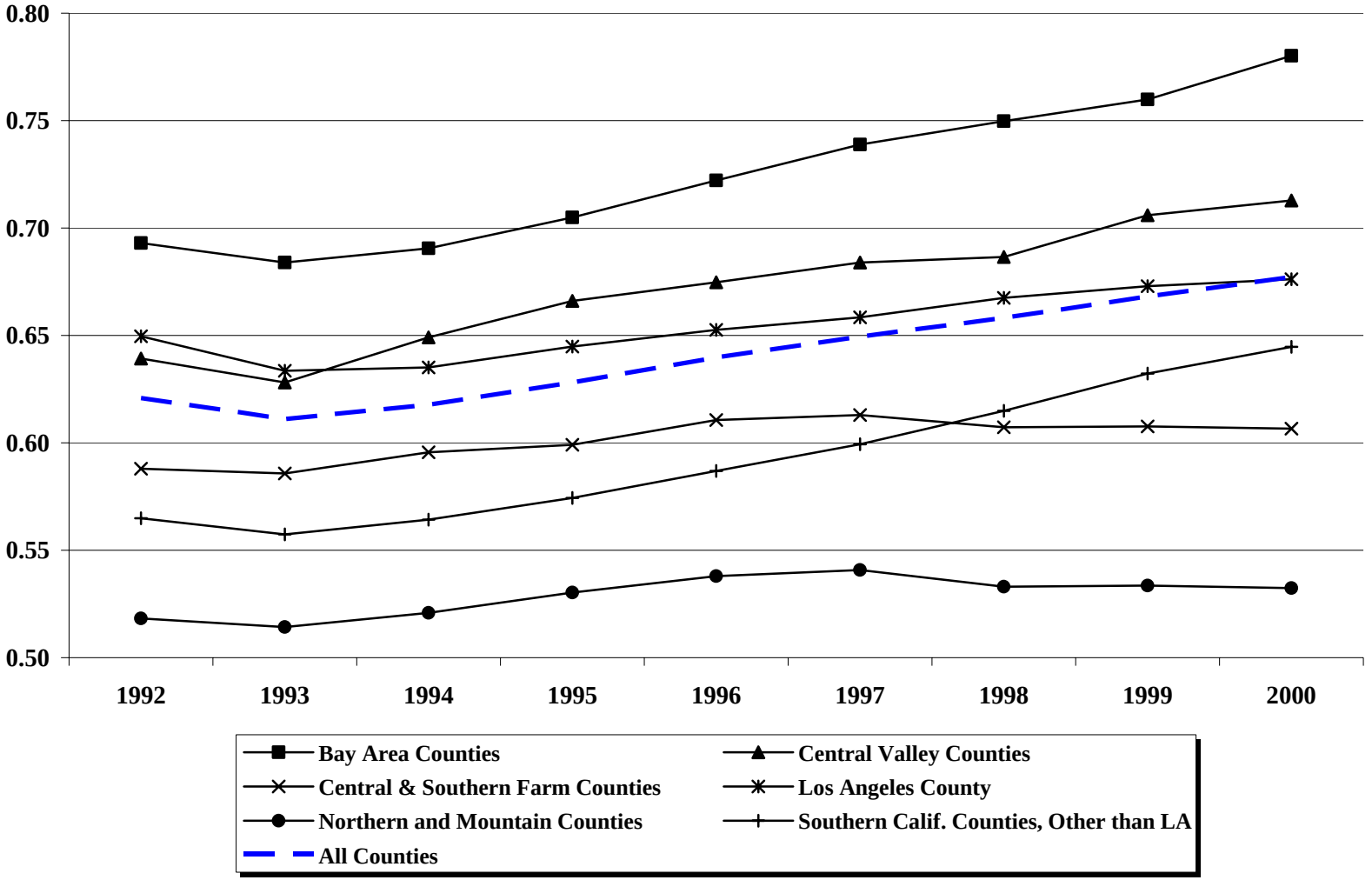


Figure 2: Annual Real Earnings per Worker in Service Sector, California, 1992-2000

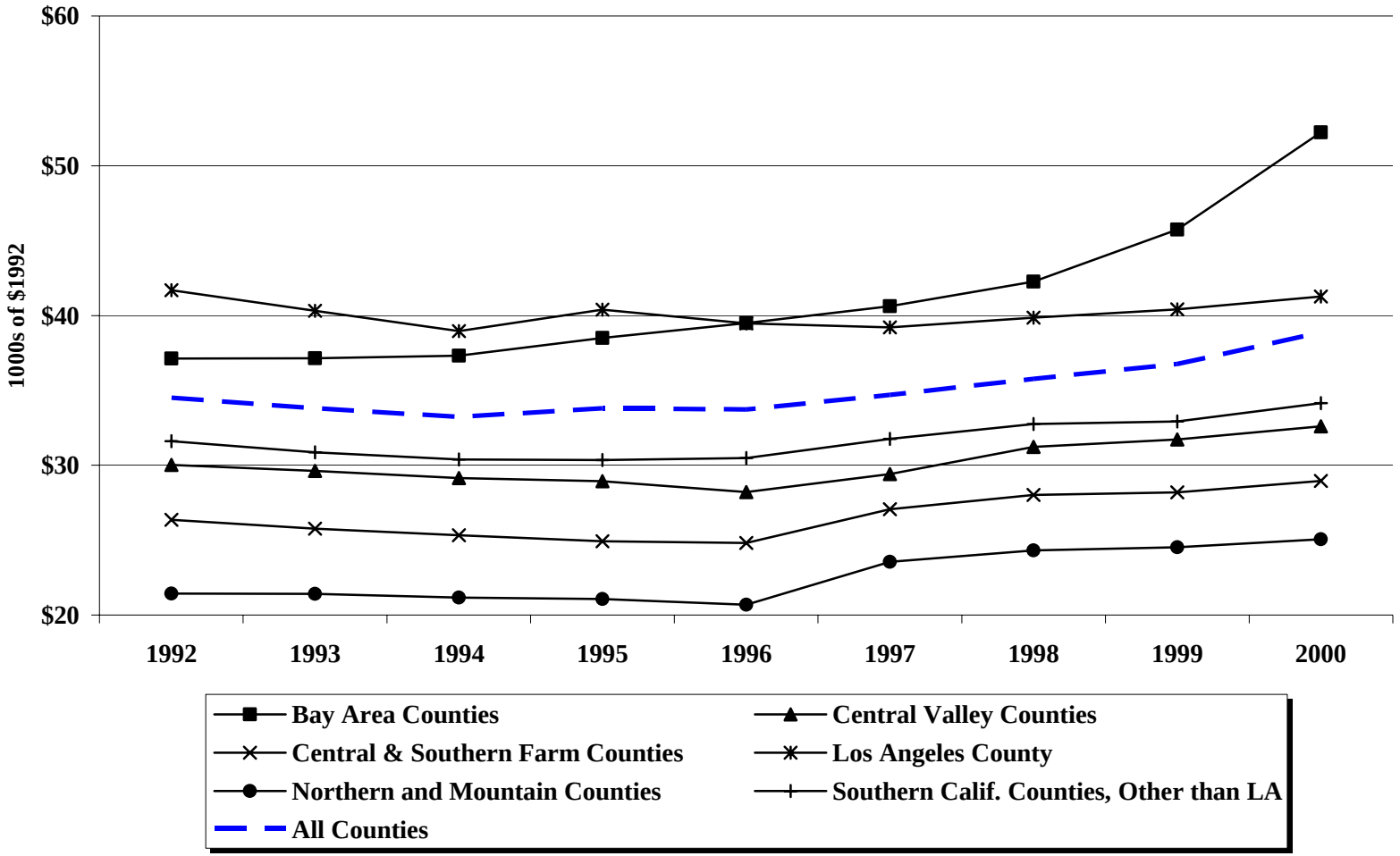


Table 1: Earned Income Tax Credit Parameters, 1987-2000 (in nominal dollars)

Year	Phase-In Rate (%)	Phase-In Range	Max Credit	Diff. in Max Credit: 2+ - 1 Child	Phase-Out Rate (%)	Phase-Out Range
1987	14.0	0-6,080	851		10.0	6,920 – 15,432
1988	14.0	0-6,240	874		10.0	9,840 – 18,576
1989	14.0	0-6,500	910		10.0	10,240 – 19,340
1990	14.0	0-6,810	953		10.0	10,730 – 20,264
1991[a]	16.7[1]	0-7,140	1,192		11.93	11,250 – 21,250
	17.3[2]		1,235	43	12.36	11,250 – 21,250
1992[a]	17.6[1]	0-7,520	1,324		12.57	11,840 – 22,370
	18.4[2]		1,384	60	13.14	11,840 – 22,370
1993[a]	18.5[1]	0-7,750	1,434		13.21	12,200 – 23,050
	19.5[2]		1,511	77	13.93	12,200 – 23,050
1994	23.6[1]	0-7,750	2,038		15.98	11,000 – 23,755
	30.0[2]	0-8,245	2,528	490	17.68	11,000 – 25,296
	7.65[3]	0-4,000	306		7.65	5,000 – 9,000
1995	34.0[1]	0-6,160	2,094		15.98	11,290 – 24,396
	36.0[2]	0-8,640	3,110	1,016	20.22	11,290 – 26,673
	7.65[3]	0-4,100	314		7.65	5,130 – 9,230
1996	34.0[1]	0-6,330	2,152		15.98	11,610 – 25,078
	40.0[2]	0-8,890	3,556	1,404	21.06	11,610 – 28,495
	7.65[3]	0-4,220	323		7.65	5,280 – 9,500
1997	34.0[1]	0-6,500	2,210		15.98	11,930 – 25,750
	40.0[2]	0-9,140	3,656	1,446	21.06	11,930 – 29,290
	7.65[3]	0-4,340	332		7.65	5,430 – 9,770
1998	34.0[1]	0-6,680	2,271		15.98	12,260 – 26,473
	40.0[2]	0-9,390	3,756	1,485	21.06	12,260 – 30,095
	7.65[3]	0-4,460	341		7.65	5,570 – 10,030
1999	34.0[1]	0-6,800	2,312		15.98	12,460 – 26,928
	40.0[2]	0-9,540	3,816	1,504	21.06	12,460 – 30,580
	7.65[3]	0-4,530	347		7.65	5,670 – 10,200
2000	34.0[1]	0-6,920	2,353		15.98	12,690 – 27,413
	40.0[2]	0-9,720	3,888	1,535	21.06	12,690 – 31,152
	7.65[3]	0-4,610	353		7.65	5,770 – 10,380

Source: *1998 Green Book*, Committee on Ways and Means, U.S. House of Representatives, U.S. Government Printing Office, page 867. 1998 through 2000 parameters come from Publication 596, Internal Revenue Service.

[a] Basic credit only. Does not include supplemental young child or health insurance credits.

[1] Taxpayers with one qualifying child.

[2] Taxpayers with more than one qualifying child.

[3] Childless taxpayers.

Table 2a: Employment Rates (in Percentages) by Family Size, 1991 – 2000, Cross-Sectional Sample

Year	All Cases	Cases with One Child	Cases with 2+ Children	Difference (2+ - One)	Diff-in-Diff (199x - 1991-93 Average)
1991	36.35	39.43	33.82	-5.61	
	[65,567]	[27,568]	[37,999]		
1992	34.82	38.35	31.76	-6.59	
	[56,787]	[24,433]	[32,354]		
1993	35.40	38.89	32.3	-6.59	
	[52,630]	[23,098]	[29,532]		
1994	40.20	43.65	37.12	-6.53	-0.31
	[54,407]	[24,005]	[30,402]		(0.66)
1995	44.89	48.44	41.64	-6.8	-0.58
	[56,664]	[25,500]	[31,164]		(0.66)
1996	48.51	51.03	45.87	-5.16	0.79
	[55,713]	[25,457]	[30,256]		(0.67)
1997	55.20	57.73	52.78	-4.95	1.27*
	[53,444]	[24,794]	[28,650]		(0.68)
1998	61.34	62.32	60.4	-1.92	4.31***
	[48,653]	[22,479]	[26,174]		(0.69)
1999	63.60	65.01	62.22	-2.79	3.43***
	[41,039]	[19,066]	[21,973]		(0.74)
2000	65.41	66.03	64.84	-1.19	5.03***
	[42,221]	[19,731]	[22,490]		(0.74)

The entries in the table give mean rates, (standard deviations), and [number of households (cases)]
*, **, and *** denote statistical significance at the 10, 5, and 1 percent levels, respectively.

Table 2b: EITC Claiming (in Percentages) by Family Size, 1991 - 2000, Cross-Sectional Sample

Year	All Cases	Cases with One Child	Cases with 2+ Children	Difference (2+ - One)	Diff-in-Diff (199x - 1991-93 Average)
1991	24.13	25.48	23.02	-2.46	
	[65,567]	[27,568]	[37,999]		
1992	22.76	23.80	21.85	-1.95	
	[56,787]	[24,433]	[32,354]		
1993	22.88	23.90	21.97	-1.93	
	[52,630]	[23,098]	[29,532]		
1994	27.84	28.87	26.91	-1.96	0.17
	[54,407]	[24,005]	[30,402]		(0.59)
1995	32.95	34.11	31.89	-2.22	-0.09
	[56,664]	[25,500]	[31,164]		(0.61)
1996	35.60	35.83	35.34	-0.49	1.69***
	[55,713]	[25,457]	[30,256]		(0.63)
1997	42.34	42.80	41.91	-0.89	1.24*
	[53,444]	[24,794]	[28,650]		(0.66)
1998	47.46	46.72	48.18	1.46	3.60***
	[48,653]	[22,479]	[26,174]		(0.69)
1999	51.28	50.18	52.35	2.17	4.31***
	[41,039]	[19,066]	[21,973]		(0.75)
2000	51.94	49.80	53.99	4.19	6.32***
	[42,221]	[19,731]	[22,490]		(0.75)

The entries in the table give mean rates, (standard deviations), and [number of cases]
*, **, and *** denote statistical significance at the 10, 5, and 1 percent levels, respectively.

Table 3: Estimates of EITC Effects on Household Employment

Variable	OLS, Cross-Sectional Sample	Household Fixed Effects, Panel Data
2+ Kids in 1994	0.0001	-0.0035
	(0.0065)	(0.0041)
2+ Kids in 1995	-0.0063	0.0072
	(0.0065)	(0.0052)
2+ Kids in 1996	0.0040	0.0123**
	(0.0066)	(0.0061)
2+ Kids in 1997	0.0092	0.0261***
	(0.0067)	(0.0071)
2+ Kids in 1998	0.0382***	0.0324***
	(0.0070)	(0.0083)
2+ Kids in 1999	0.0278***	0.0341***
	(0.0076)	(0.0100)
2+ Kids in 2000	0.0425***	0.0295**
	(0.0075)	(0.0123)
2 Kids in Household	0.0337***	-0.0102
	(0.0055)	(0.0151)
3 Kids in Household	0.0606***	-0.0042
	(0.0094)	(0.0275)
4 Kids in Household	0.0799***	0.0135
	(0.0137)	(0.0393)
5 Kids in Household	0.0897***	0.0413
	(0.0182)	(0.0508)
6 Kids in Household	0.1132***	0.0859
	(0.0232)	(0.0616)
7 Kids in Household	0.1607***	0.1293*
	(0.0290)	(0.0734)
8 Kids in Household	0.2186***	0.1801**
	(0.0369)	(0.0876)
9 or more Kids in Household	0.2477***	0.2230**
	(0.0475)	(0.1098)
No. of Kids Age 1	-0.1600***	-0.1206***
	(0.0051)	(0.0116)
No. of Kids Age 2	-0.1330***	-0.0888***
	(0.0050)	(0.0110)
No. of Kids Age 3	-0.1127***	-0.0653***
	(0.0050)	(0.0104)
No. of Kids Age 4	-0.0926***	-0.0455***
	(0.0050)	(0.0099)
No. of Kids Age 5	-0.0790***	-0.0384***
	(0.0050)	(0.0094)

Table 3 (Continued)

Variable	OLS, Cross-Sectional Sample	Household Fixed Effects, Panel Data
No. of Kids Age 6	-0.0685***	-0.0320***
	(0.0050)	(0.0090)
No. of Kids Age 7	-0.0538***	-0.0245***
	(0.0050)	(0.0086)
No. of Kids Age 8	-0.0496***	-0.0239***
	(0.0051)	(0.0082)
No. of Kids Age 9	-0.0463***	-0.0211***
	(0.0051)	(0.0079)
No. of Kids Age 10	-0.0431***	-0.0201***
	(0.0052)	(0.0076)
No. of Kids Age 11	-0.0335***	-0.0162**
	(0.0052)	(0.0073)
No. of Kids Age 12	-0.0319***	-0.0136**
	(0.0053)	(0.0069)
No. of Kids Age 13	-0.0224***	-0.0131**
	(0.0054)	(0.0066)
No. of Kids Age 14	-0.0209***	-0.0072
	(0.0055)	(0.0063)
No. of Kids Age 15	-0.0214***	-0.0065
	(0.0057)	(0.0060)
No. of Kids Age 16	-0.0236***	-0.0080
	(0.0058)	(0.0057)
No. of Kids Age 17	-0.0166***	-0.0101**
No. of Observations	527,125	1,637,855
P-Value for Test of 2+ Kids in 1994-2000 = 0	0.0000	0.0004

Table 4: Estimates of EITC Effects on Whether Household Claimed the EITC on Tax Return

Variable	OLS, Cross-Sectional Sample	Household Fixed Effects, Panel Data
2+ Kids in 1994	0.0041	0.0006
	(0.0059)	(0.0038)
2+ Kids in 1995	-0.0032	0.0065
	(0.0061)	(0.0049)
2+ Kids in 1996	0.0103*	0.0184***
	(0.0062)	(0.0058)
2+ Kids in 1997	0.0050	0.0190***
	(0.0066)	(0.0068)
2+ Kids in 1998	0.0249***	0.0170**
	(0.0070)	(0.0081)
2+ Kids in 1999	0.0283***	0.0233**
	(0.0077)	(0.0098)
2+ Kids in 2000	0.0448***	0.0194
	(0.0076)	(0.0122)
2 Kids in Household	0.0351***	-0.0601***
	(0.0051)	(0.0143)
3 Kids in Household	0.0532***	-0.1193***
	(0.0089)	(0.0261)
4 Kids in Household	0.0593***	-0.1736***
	(0.0130)	(0.0373)
5 Kids in Household	0.0591***	-0.2188***
	(0.0175)	(0.0484)
6 Kids in Household	0.0728***	-0.2535***
	(0.0222)	(0.0587)
7 Kids in Household	0.1093***	-0.3096***
	(0.0278)	(0.0697)
8 Kids in Household	0.1197***	-0.2874***
	(0.0351)	(0.0833)
9 or more Kids in Household	0.1830***	-0.2442**
	(0.0448)	(0.1044)
No. of Kids Age 1	-0.1153***	-0.0093
	(0.0049)	(0.0110)
No. of Kids Age 2	-0.0971***	0.0072
	(0.0048)	(0.0104)
No. of Kids Age 3	-0.0762***	0.0200**
	(0.0048)	(0.0099)
No. of Kids Age 4	-0.0671***	0.0280***
	(0.0047)	(0.0094)
No. of Kids Age 5	-0.0541***	0.0294***
	(0.0048)	(0.0090)

Table 4 (Continued)

Variable	OLS, Cross-Sectional Sample	Household Fixed Effects, Panel Data
No. of Kids Age 6	-0.0434***	0.0308***
	(0.0048)	(0.0086)
No. of Kids Age 7	-0.0340***	0.0330***
	(0.0048)	(0.0082)
No. of Kids Age 8	-0.0316***	0.0316***
	(0.0049)	(0.0079)
No. of Kids Age 9	-0.0287***	0.0312***
	(0.0049)	(0.0076)
No. of Kids Age 10	-0.0241***	0.0301***
	(0.0050)	(0.0073)
No. of Kids Age 11	-0.0200***	0.0298***
	(0.0050)	(0.0070)
No. of Kids Age 12	-0.0212***	0.0300***
	(0.0051)	(0.0067)
No. of Kids Age 13	-0.0103**	0.0270***
	(0.0052)	(0.0064)
No. of Kids Age 14	-0.0109**	0.0251***
	(0.0053)	(0.0061)
No. of Kids Age 15	-0.0163***	0.0205***
	(0.0054)	(0.0058)
No. of Kids Age 16	-0.0152***	0.0147***
	(0.0055)	(0.0055)
No. of Kids Age 17	-0.0037	0.0088*
	(0.0055)	(0.0049)
No. of Observations	527,125	1,637,855
P-Value for Test of 2+ Kids in 1994-2000 = 0	0.0000	0.0353

Table 5A: Assessing the Validity of Strategy for Identifying EITC Effects on Household Employment [Family Fixed Effects Estimation on Panel Data]

Variable	All Households		Households that Did Not File Tax Return
	(1)	(2)	(3)
2+ Kids in 1994	-0.0035	-0.0016	-0.0020
	(0.0041)	(0.0051)	(0.0065)
2+ Kids in 1995	0.0072	0.0043	0.0028
	(0.0052)	(0.0067)	(0.0086)
2+ Kids in 1996	0.0123**	0.0114	-0.0016
	(0.0061)	(0.0082)	(0.0107)
2+ Kids in 1997	0.0261***	0.0289***	0.0113
	(0.0071)	(0.0098)	(0.0131)
2+ Kids in 1998	0.0324***	0.0331***	0.0048
	(0.0083)	(0.0116)	(0.0161)
2+ Kids in 1999	0.0341***	0.0368***	0.0037
	(0.0100)	(0.0137)	(0.0200)
2+ Kids in 2000	0.0295**	0.0291*	-0.0009
	(0.0123)	(0.0168)	(0.0257)
[2+ Kids – 3+ Kids] in 1994		-0.0030	-0.0009
		(0.0051)	(0.0062)
[2+ Kids – 3+ Kids] in 1995		0.0055	-0.0010
		(0.0063)	(0.0080)
[2+ Kids – 3+ Kids] in 1996		0.0021	-0.0014
		(0.0074)	(0.0096)
[2+ Kids – 3+ Kids] in 1997		-0.0042	-0.0012
		(0.0086)	(0.0115)
[2+ Kids – 3+ Kids] in 1998		-0.0004	0.0005
		(0.0100)	(0.0141)
[2+ Kids – 3+ Kids] in 1999		-0.0039	-0.0071
		(0.0119)	(0.0181)
[2+ Kids – 3+ Kids] in 2000		0.0015	0.0077
		(0.0149)	(0.0244)
No. of Observations			
P-Value for Test of 2+ Kids in 1994-2000 = 0	0.0004	0.0112	0.8423
P-Value for Test of [2+ Kids – 3+ Kids] in 1994-2000 = 0		0.7389	0.9989

Table 5B: Assessing the Validity of Strategy for Identifying EITC Effects on Whether Household Claimed the EITC on Tax Return [Family Fixed Effects Estimation on Panel Data]

Variable	(1)	(2)
2+ Kids in 1994	0.0006	0.0057
	(0.0038)	(0.0048)
2+ Kids in 1995	0.0065	0.0131**
	(0.0049)	(0.0064)
2+ Kids in 1996	0.0184***	0.0233***
	(0.0058)	(0.0078)
2+ Kids in 1997	0.0190***	0.0279***
	(0.0068)	(0.0094)
2+ Kids in 1998	0.0170**	0.0170
	(0.0081)	(0.0112)
2+ Kids in 1999	0.0233**	0.0344**
	(0.0098)	(0.0134)
2+ Kids in 2000	0.0194	0.0255
	(0.0122)	(0.0165)
[2+ Kids – 3+ Kids] in 1994		-0.0078*
		(0.0047)
[2+ Kids – 3+ Kids] in 1995		-0.0100*
		(0.0059)
[2+ Kids – 3+ Kids] in 1996		-0.0067
		(0.0070)
[2+ Kids – 3+ Kids] in 1997		-0.0131
		(0.0082)
[2+ Kids – 3+ Kids] in 1998		0.0033
		(0.0097)
[2+ Kids – 3+ Kids] in 1999		-0.0158
		(0.0119)
[2+ Kids – 3+ Kids] in 2000		-0.0063
		(0.0149)
P-Value for Test of 2+ Kids in 1994-2000 = 0	0.0353	0.0174
P-Value for Test of [2+ Kids – 3+ Kids] in 1994-2000 = 0		0.0669

Table 5C: Estimates of EITC Effects on Household Employment Using 1992 as Base Year [Family Fixed Effects Estimation on Panel Data]

Variable	(1)	(1)
2+ Kids in 1991	-0.0040	-0.0040
	(0.0043)	(0.0053)
2+ Kids in 1993	0.0008	-0.0001
	(0.0042)	(0.0051)
2+ Kids in 1994	-0.0031	-0.0012
	(0.0052)	(0.0068)
2+ Kids in 1995	0.0079	0.0054
	(0.0061)	(0.0084)
2+ Kids in 1996	0.0132*	0.0130
	(0.0070)	(0.0100)
2+ Kids in 1997	0.0272***	0.0310***
	(0.0079)	(0.0117)
2+ Kids in 1998	0.0337***	0.0357***
	(0.0091)	(0.0136)
2+ Kids in 1999	0.0357***	0.0400**
	(0.0108)	(0.0159)
2+ Kids in 2000	0.0314**	0.0330*
	(0.0130)	(0.0191)
[2+ Kids – 3+ Kids] in 1991		-0.0005
		(0.0052)
[2+ Kids – 3+ Kids] in 1993		0.0018
		(0.0051)
[2+ Kids – 3+ Kids] in 1994		-0.0027
		(0.0065)
[2+ Kids – 3+ Kids] in 1995		0.0056
		(0.0076)
[2+ Kids – 3+ Kids] in 1996		0.0018
		(0.0088)
[2+ Kids – 3+ Kids] in 1997		-0.0049
		(0.0099)
[2+ Kids – 3+ Kids] in 1998		-0.0015
		(0.0114)
[2+ Kids – 3+ Kids] in 1999		-0.0054
		(0.0133)
[2+ Kids – 3+ Kids] in 2000		-0.0004
		(0.0163)
P-Value for Test of 2+ Kids 1994-2000 = 0	0.0003	0.0088
P-Value for Test of 2+ Kids in 1991 & 1993 = 0	0.6069	
P-Value for Test of [2+ Kids–3+ Kids] for 1994-2000 = 0		0.7373

Table 5D: Estimates of EITC Effects on Household Employment Using Alternative Earnings Thresholds [Family Fixed Effects Estimation on Panel Data]

Variable	Household Employment = 1 if Earnings > $500		Household Employment = 1 if Earnings > $1,000	
	(1)	(2)	(3)	(4)
2+ Kids in 1994	-0.0015	-0.0007	-0.0039	-0.0030
	(0.0040)	(0.0049)	(0.0039)	(0.0047)
2+ Kids in 1995	0.0063	0.0032	0.0037	0.0026
	(0.0050)	(0.0065)	(0.0049)	(0.0063)
2+ Kids in 1996	0.0130**	0.0090	0.0101*	0.0067
	(0.0060)	(0.0079)	(0.0059)	(0.0077)
2+ Kids in 1997	0.0217***	0.0216**	0.0176***	0.0159*
	(0.0070)	(0.0095)	(0.0068)	(0.0092)
2+ Kids in 1998	0.0331***	0.0268**	0.0264***	0.0225**
	(0.0082)	(0.0113)	(0.0081)	(0.0110)
2+ Kids in 1999	0.0308***	0.0241*	0.0201**	0.0141
	(0.0100)	(0.0134)	(0.0098)	(0.0131)
2+ Kids in 2000	0.0259**	0.0195	0.0141	0.0065
	(0.0123)	(0.0164)	(0.0121)	(0.0162)
[2+ Kids – 3+ Kids] in 1994		-0.0025		-0.0027
		(0.0049)		(0.0047)
[2+ Kids – 3+ Kids] in 1995		0.0037		0.0002
		(0.0061)		(0.0059)
[2+ Kids – 3+ Kids] in 1996		0.0045		0.0035
		(0.0072)		(0.0070)
[2+ Kids – 3+ Kids] in 1997		-0.0031		-0.0002
		(0.0083)		(0.0081)
[2+ Kids – 3+ Kids] in 1998		0.0073		0.0033
		(0.0097)		(0.0095)
[2+ Kids – 3+ Kids] in 1999		0.0076		0.0064
		(0.0118)		(0.0116)
[2+ Kids – 3+ Kids] in 2000		0.0063		0.0086
		(0.0147)		(0.0146)
P-Value for Test of 2+ Kids for 1994-2000 = 0	0.0026	0.1482	0.0083	0.1636
P-Value for Test of [2+ Kids – 3+ Kids] for 1994-2000 = 0		0.6594		0.9676

Appendix Table 1a: Weighted Descriptive Statistics for the Cross Sectional Analysis Sample (*N*=527,125)

Variable		Mean	Std Dev	Min	Max
Outcome variables of interest:					
Emp	Household employed in year (= 1if at least 1 adult employed; = 0 otherwise)	0.477	0.499	0	1
FiledTax	Household filed tax return in year (= 1 yes; = 0 no)	0.425	0.494	0	1
ClaimEITC	Household filed tax return and claimed EITC in year (= 1, yes; = 0, no)	0.351	0.477	0	1
Calendar Year Indicators:					
Year_1991	Year = 1991	0.120	0.325	0	1
Year_1992	Year = 1992	0.106	0.307	0	1
Year_1993	Year = 1993	0.098	0.297	0	1
Year_1994	Year = 1994	0.102	0.302	0	1
Year_1995	Year = 1995	0.105	0.307	0	1
Year_1996	Year = 1996	0.105	0.307	0	1
Year_1997	Year = 1997	0.102	0.302	0	1
Year_1998	Year = 1998	0.093	0.290	0	1
Year_1999	Year = 1999	0.079	0.270	0	1
Year_2000	Year = 2000	0.091	0.288	0	1
Treatment and "Testing" Variables:					
2+ Kids in 1991	Household had 2 or more Kids in 1991	0.066	0.248	0	1
2+ Kids in 1992	Household had 2 or more Kids in 1992	0.057	0.231	0	1
2+ Kids in 1993	Household had 2 or more Kids in 1993	0.052	0.221	0	1
2+ Kids in 1994	Household had 2 or more Kids in 1994	0.054	0.225	0	1
2+ Kids in 1995	Household had 2 or more Kids in 1995	0.055	0.228	0	1
2+ Kids in 1996	Household had 2 or more Kids in 1996	0.054	0.226	0	1
2+ Kids in 1997	Household had 2 or more Kids in 1997	0.052	0.222	0	1
2+ Kids in 1998	Household had 2 or more Kids in 1998	0.047	0.211	0	1
2+ Kids in 1999	Household had 2 or more Kids in 1999	0.040	0.196	0	1
2+ Kids in 2000	Household had 2 or more Kids in 2000	0.046	0.211	0	1
[2+ Kids–3+ Kids] in 1991	Household had exactly two children in 1991	0.038	0.191	0	1
[2+ Kids–3+ Kids] in 1992	Household had exactly two children in 1992	0.032	0.176	0	1
[2+ Kids–3+ Kids] in 1993	Household had exactly two children in 1993	0.029	0.169	0	1
[2+ Kids–3+ Kids] in 1994	Household had exactly two children in 1994	0.031	0.172	0	1
[2+ Kids–3+ Kids] in 1995	Household had exactly two children in 1995	0.032	0.176	0	1
[2+ Kids–3+ Kids] in 1996	Household had exactly two children in 1996	0.031	0.174	0	1
[2+ Kids–3+ Kids] in 1997	Household had exactly two children in 1997	0.030	0.171	0	1
[2+ Kids–3+ Kids] in 1998	Household had exactly two children in 1998	0.027	0.162	0	1
[2+ Kids–3+ Kids] in 1999	Household had exactly two children in 1999	0.023	0.150	0	1
[2+ Kids–3+ Kids] in 2000	Household had exactly two children in 2000	0.027	0.161	0	1
Number and Ages of Children:					
1 Kid in Household	Household has 1 child (= 1, yes; = 0, no)	0.478	0.499	0	1
2 Kids in Household	Household has 2 children (= 1, yes; = 0, no)	0.300	0.458	0	1
3 Kids in Household	Household has 3 children (= 1, yes; = 0, no)	0.142	0.349	0	1
4 Kids in Household	Household has 4 children (= 1, yes; = 0, no)	0.053	0.225	0	1
5 Kids in Household	Household has 5 children (= 1, yes; = 0, no)	0.018	0.132	0	1
6 Kids in Household	Household has 6 children (= 1, yes; = 0, no)	0.006	0.077	0	1
7 Kids in Household	Household has 7 children (= 1, yes; = 0, no)	0.002	0.046	0	1
8 Kids in Household	Household has 8 children (= 1, yes; = 0, no)	0.001	0.030	0	1
9 or more Kids in Household	Household has 9 or more children (= 1, yes; = 0, no)	0.000	0.021	0	1
No. of Kids Age 0	Number of Kids Age 0 in Household	0.000	0.000	0	0
No. of Kids Age 1	Number of Kids Age 1 in Household	0.142	0.358	0	4
No. of Kids Age 2	Number of Kids Age 2 in Household	0.164	0.379	0	5
No. of Kids Age 3	Number of Kids Age 3 in Household	0.163	0.376	0	3
No. of Kids Age 4	Number of Kids Age 4 in Household	0.157	0.371	0	4
No. of Kids Age 5	Number of Kids Age 5 in Household	0.142	0.355	0	3
No. of Kids Age 6	Number of Kids Age 6 in Household	0.131	0.343	0	3
No. of Kids Age 7	Number of Kids Age 7 in Household	0.119	0.329	0	4

Variable		Mean	Std Dev	Min	Max
No. of Kids Age 8	Number of Kids Age 8 in Household	0.110	0.318	0	3
No. of Kids Age 9	Number of Kids Age 9 in Household	0.101	0.307	0	5
No. of Kids Age 10	Number of Kids Age 10 in Household	0.094	0.296	0	4
No. of Kids Age 11	Number of Kids Age 11 in Household	0.087	0.285	0	3
No. of Kids Age 12	Number of Kids Age 12 in Household	0.081	0.276	0	3
No. of Kids Age 13	Number of Kids Age 13 in Household	0.075	0.267	0	3
No. of Kids Age 14	Number of Kids Age 14 in Household	0.070	0.259	0	3
No. of Kids Age 15	Number of Kids Age 15 in Household	0.065	0.250	0	4
No. of Kids Age 16	Number of Kids Age 16 in Household	0.061	0.244	0	3
No. of Kids Age 17	Number of Kids Age 17 in Household	0.055	0.232	0	3
No. of Kids Age 18	Number of Kids Age 18 in Household	0.048	0.217	0	3
Demographic Characteristics of Households:					
Ave. Adult Age	Mean Age of Adults in Household in given year	31.374	8.507	20	103
Female Head	Households is Headed by Female (= 1, yes; = 0, no)	0.933	0.249	0	1
Hispanic	All adults in Household are Hispanic	0.284	0.451	0	1
Black	All adults in Household are black	0.224	0.417	0	1
Asian	All adults in Households are Asian	0.072	0.259	0	1
Other/Mixed Race/Ethnicity	Adults in Household are of other and/or mixed races	0.013	0.113	0	1
County Welfare Variables and Year of Welfare Entry by Year:					
GAIN Proportion	Average Annual Proportion AFDC/TANF recipients in County of Residence in GAIN/WTW program	0.362	0.262	0.036	1
Child Care Expend.	Per capita public Expenditures (in 1000s of $1992/93) on Child Care for Low-Income Population in County of Residence county in given year	0.023	0.039	0	0.174
Entered AFDC/TANF in 1986	Household first went on AFDC/TANF in 1986 (= 1, yes; = 0, no)	0.070	0.255	0	1
Entered AFDC/TANF in 1987	Household first went on AFDC/TANF in 1987 (= 1, yes; = 0, no)	0.021	0.142	0	1
Entered AFDC/TANF in 1988	Household first went on AFDC/TANF in 1988 (= 1, yes; = 0, no)	0.032	0.177	0	1
Entered AFDC/TANF in 1989	Household first went on AFDC/TANF in 1989 (= 1, yes; = 0, no)	0.050	0.218	0	1
Entered AFDC/TANF in 1990	Household first went on AFDC/TANF in 1990 (= 1, yes; = 0, no)	0.092	0.288	0	1
Entered AFDC/TANF in 1991	Household first went on AFDC/TANF in 1991 (= 1, yes; = 0, no)	0.103	0.303	0	1
Entered AFDC/TANF in 1992	Household first went on AFDC/TANF in 1992 (= 1, yes; = 0, no)	0.090	0.286	0	1
Entered AFDC/TANF in 1993	Household first went on AFDC/TANF in 1993 (= 1, yes; = 0, no)	0.101	0.301	0	1
Entered AFDC/TANF in 1994	Household first went on AFDC/TANF in 1994 (= 1, yes; = 0, no)	0.099	0.299	0	1
Entered AFDC/TANF in 1995	Household first went on AFDC/TANF in 1995 (= 1, yes; = 0, no)	0.089	0.285	0	1
Entered AFDC/TANF in 1996	Household first went on AFDC/TANF in 1996 (= 1, yes; = 0, no)	0.077	0.267	0	1
Entered AFDC/TANF in 1997	Household first went on AFDC/TANF in 1997 (= 1, yes; = 0, no)	0.068	0.251	0	1
Entered AFDC/TANF in 1998	Household first went on AFDC/TANF in 1998 (= 1, yes; = 0, no)	0.056	0.230	0	1
Entered AFDC/TANF in 1999	Household first went on AFDC/TANF in 1999 (= 1, yes; = 0, no)	0.052	0.223	0	1
Local Labor Market Characteristics:					
Emp-to-Pop	Annual Employment-to-Population ratio in County of Residence	0.622	0.107	0.317	1.330
Prop. Employ. in Manufacturing	Proportion of County of Residence Employment in Manufacturing in given year	0.134	0.051	0	0.305
Prop. Employ. in Services	Proportion of County of Residence Employment in Services in given year	0.303	0.074	0	0.931
Prop. Employ. in Retail Trade	Proportion of County of Residence Employment in Retail Trade in given year	0.176	0.026	0.025	0.399
Ave. Earnings, All Sectors	Average Annual Earnings per Worker in County of Residence in 1000s of 1992/93$, All Sectors	34.417	6.690	14.278	78.381
Ave. Earnings, Manufacturing	Average Annual Earnings per Worker in County of Residence in 1000s of 1992/93$, Manufacturing Sector	41.515	9.556	0	136.017
Ave. Earnings, Services	Average Annual Earnings per Worker in County of Residence in 1000s of 1992/93$, Service Sector	33.313	6.979	0	81.233
Ave. Earnings, Retail Trade	Average Annual Earnings per Worker in County of Residence in 1000s of 1992/93$, Retail Trade Sector	19.681	2.310	8.705	33.310
County of Residence Dummy Variables:					
Alameda	Resided in Alameda County at sampling	0.043	0.203	0	1
Alpine	Resided in Alpine County at sampling	0.000	0.009	0	1
Amador	Resided in Amador County at sampling	0.001	0.029	0	1
Butte	Resided in Butte County at sampling	0.009	0.093	0	1
Calaveras	Resided in Calaveras County at sampling	0.002	0.039	0	1

Variable		Mean	Std Dev	Min	Max
Colusa	Resided in Colusa County at sampling	0.001	0.024	0	1
Contra Costa	Resided in Contra Costa County at sampling	0.022	0.148	0	1
Del Norte	Resided in Del Norte County at sampling	0.002	0.041	0	1
El Dorado	Resided in El Dorado County at sampling	0.004	0.062	0	1
Fresno	Resided in Fresno County at sampling	0.034	0.180	0	1
Glenn	Resided in Glenn County at sampling	0.001	0.033	0	1
Humboldt	Resided in Humboldt County at sampling	0.006	0.076	0	1
Imperial	Resided in Imperial County at sampling	0.006	0.080	0	1
Inyo	Resided in Inyo County at sampling	0.001	0.027	0	1
Kern	Resided in Kern County at sampling	0.029	0.167	0	1
Kings	Resided in Kings County at sampling	0.006	0.079	0	1
Lake	Resided in Lake County at sampling	0.004	0.060	0	1
Lassen	Resided in Lassen County at sampling	0.001	0.037	0	1
Los Angeles	Resided in Los Angeles County at sampling	0.285	0.451	0	1
Madera	Resided in Madera County at sampling	0.004	0.066	0	1
Marin	Resided in Marin County at sampling	0.002	0.049	0	1
Mariposa	Resided in Mariposa County at sampling	0.001	0.027	0	1
Mendocino	Resided in Mendocino County at sampling	0.004	0.064	0	1
Merced	Resided in Merced County at sampling	0.010	0.100	0	1
Modoc	Resided in Modoc County at sampling	0.001	0.024	0	1
Mono	Resided in Mono County at sampling	0.000	0.014	0	1
Monterey	Resided in Monterey County at sampling	0.012	0.110	0	1
Napa	Resided in Napa County at sampling	0.003	0.050	0	1
Nevada	Resided in Nevada County at sampling	0.002	0.048	0	1
Orange	Resided in Orange County at sampling	0.041	0.199	0	1
Placer	Resided in Placer County at sampling	0.006	0.075	0	1
Plumas	Resided in Plumas County at sampling	0.001	0.029	0	1
Riverside	Resided in Riverside County at sampling	0.050	0.219	0	1
Sacramento	Resided in Sacramento County at sampling	0.052	0.222	0	1
San Benito	Resided in San Benito County at sampling	0.001	0.037	0	1
San Bernardino	Resided in San Bernardino County at sampling	0.083	0.275	0	1
San Diego	Resided in San Diego County at sampling	0.073	0.260	0	1
San Francisco	Resided in San Francisco County at sampling	0.014	0.117	0	1
San Joaquin	Resided in San Joaquin County at sampling	0.024	0.154	0	1
San Luis Obispo	Resided in San Luis Obispo County at sampling	0.005	0.072	0	1
San Mateo	Resided in San Mateo County at sampling	0.008	0.089	0	1
Santa Barbara	Resided in Santa Barbara County at sampling	0.008	0.088	0	1
Santa Clara	Resided in Santa Clara County at sampling	0.033	0.179	0	1
Santa Cruz	Resided in Santa Cruz County at sampling	0.005	0.073	0	1
Shasta	Resided in Shasta County at sampling	0.008	0.091	0	1
Sierra	Resided in Sierra County at sampling	0.000	0.012	0	1
Siskiyou	Resided in Siskiyou County at sampling	0.003	0.054	0	1
Solano	Resided in Solano County at sampling	0.013	0.113	0	1
Sonoma	Resided in Sonoma County at sampling	0.009	0.094	0	1
Stanislaus	Resided in Stanislaus County at sampling	0.018	0.134	0	1
Sutter	Resided in Sutter County at sampling	0.003	0.058	0	1
Tehama	Resided in Tehama County at sampling	0.003	0.053	0	1
Trinity	Resided in Trinity County at sampling	0.001	0.027	0	1
Tulare	Resided in Tulare County at sampling	0.016	0.126	0	1
Tuolumne	Resided in Tuolumne County at sampling	0.002	0.047	0	1
Ventura	Resided in Ventura County at sampling	0.013	0.115	0	1
Yolo	Resided in Yolo County at sampling	0.006	0.075	0	1
Yuba	Resided in Yuba County at sampling	0.004	0.064	0	1

Appendix Table 1b: Weighted Descriptive Statistics for the Longitudinal Analysis Sample (N=1,637,855)

Variable		Mean	Std Dev	Min	Max
Outcome variables of interest:					
Emp	Household employed in year (= 1if at least 1 adult employed; = 0 otherwise)	0.445	0.497	0	1
FiledTax	Household filed tax return in year (= 1 yes; = 0 no)	0.390	0.488	0	1
ClaimEITC	Household filed tax return and claimed EITC in year (= 1, yes; = 0, no)	0.317	0.465	0	1
Calendar Year Indicators:					
Year_1991	Year = 1991	0.128	0.334	0	1
Year_1992	Year = 1992	0.122	0.328	0	1
Year_1993	Year = 1993	0.122	0.327	0	1
Year_1994	Year = 1994	0.123	0.328	0	1
Year_1995	Year = 1995	0.120	0.325	0	1
Year_1996	Year = 1996	0.113	0.316	0	1
Year_1997	Year = 1997	0.107	0.309	0	1
Year_1998	Year = 1998	0.081	0.272	0	1
Year_1999	Year = 1999	0.055	0.228	0	1
Year_2000	Year = 2000	0.029	0.169	0	1
Treatment and "Testing" Variables:					
2+ Kids in 1991	Household had 2 or more Kids in 1991	0.069	0.254	0	1
2+ Kids in 1992	Household had 2 or more Kids in 1992	0.066	0.248	0	1
2+ Kids in 1993	Household had 2 or more Kids in 1993	0.065	0.246	0	1
2+ Kids in 1994	Household had 2 or more Kids in 1994	0.065	0.247	0	1
2+ Kids in 1995	Household had 2 or more Kids in 1995	0.063	0.243	0	1
2+ Kids in 1996	Household had 2 or more Kids in 1996	0.059	0.235	0	1
2+ Kids in 1997	Household had 2 or more Kids in 1997	0.056	0.230	0	1
2+ Kids in 1998	Household had 2 or more Kids in 1998	0.042	0.200	0	1
2+ Kids in 1999	Household had 2 or more Kids in 1999	0.028	0.165	0	1
2+ Kids in 2000	Household had 2 or more Kids in 2000	0.015	0.121	0	1
[2+ Kids–3+ Kids] in 1991	Household had exactly two children in 1991	0.039	0.194	0	1
[2+ Kids–3+ Kids] in 1992	Household had exactly two children in 1992	0.037	0.190	0	1
[2+ Kids–3+ Kids] in 1993	Household had exactly two children in 1993	0.037	0.189	0	1
[2+ Kids–3+ Kids] in 1994	Household had exactly two children in 1994	0.037	0.190	0	1
[2+ Kids–3+ Kids] in 1995	Household had exactly two children in 1995	0.036	0.187	0	1
[2+ Kids–3+ Kids] in 1996	Household had exactly two children in 1996	0.034	0.180	0	1
[2+ Kids–3+ Kids] in 1997	Household had exactly two children in 1997	0.032	0.175	0	1
[2+ Kids–3+ Kids] in 1998	Household had exactly two children in 1998	0.024	0.152	0	1
[2+ Kids–3+ Kids] in 1999	Household had exactly two children in 1999	0.016	0.125	0	1
[2+ Kids–3+ Kids] in 2000	Household had exactly two children in 2000	0.009	0.092	0	1
Number and Ages of Children:					
1 Kid in Household	Household has 1 child (= 1, yes; = 0, no)	0.478	0.500	0	1
2 Kids in Household	Household has 2 children (= 1, yes; = 0, no)	0.301	0.459	0	1
3 Kids in Household	Household has 3 children (= 1, yes; = 0, no)	0.142	0.349	0	1
4 Kids in Household	Household has 4 children (= 1, yes; = 0, no)	0.053	0.224	0	1
5 Kids in Household	Household has 5 children (= 1, yes; = 0, no)	0.017	0.131	0	1
6 Kids in Household	Household has 6 children (= 1, yes; = 0, no)	0.006	0.076	0	1
7 Kids in Household	Household has 7 children (= 1, yes; = 0, no)	0.002	0.045	0	1
8 Kids in Household	Household has 8 children (= 1, yes; = 0, no)	0.001	0.030	0	1
9 or more Kids in Household	Household has 9 or more children (= 1, yes; = 0, no)	0.000	0.019	0	1
No. of Kids Age 0	Number of Kids Age 0 in Household	0.105	0.313	0	5
No. of Kids Age 1	Number of Kids Age 1 in Household	0.159	0.374	0	5
No. of Kids Age 2	Number of Kids Age 2 in Household	0.165	0.379	0	5
No. of Kids Age 3	Number of Kids Age 3 in Household	0.160	0.373	0	4
No. of Kids Age 4	Number of Kids Age 4 in Household	0.149	0.363	0	4
No. of Kids Age 5	Number of Kids Age 5 in Household	0.137	0.349	0	4
No. of Kids Age 6	Number of Kids Age 6 in Household	0.125	0.336	0	4
No. of Kids Age 7	Number of Kids Age 7 in Household	0.114	0.323	0	4

Variable		Mean	Std Dev	Min	Max
No. of Kids Age 8	Number of Kids Age 8 in Household	0.105	0.312	0	5
No. of Kids Age 9	Number of Kids Age 9 in Household	0.097	0.301	0	5
No. of Kids Age 10	Number of Kids Age 10 in Household	0.090	0.291	0	4
No. of Kids Age 11	Number of Kids Age 11 in Household	0.084	0.281	0	3
No. of Kids Age 12	Number of Kids Age 12 in Household	0.078	0.273	0	3
No. of Kids Age 13	Number of Kids Age 13 in Household	0.073	0.264	0	4
No. of Kids Age 14	Number of Kids Age 14 in Household	0.067	0.255	0	4
No. of Kids Age 15	Number of Kids Age 15 in Household	0.061	0.244	0	4
No. of Kids Age 16	Number of Kids Age 16 in Household	0.047	0.214	0	3
No. of Kids Age 17	Number of Kids Age 17 in Household	0.031	0.176	0	3
No. of Kids Age 18	Number of Kids Age 18 in Household	0.015	0.125	0	3
County Welfare Variables and Year of Welfare Entry by Year:					
GAIN Proportion	Average Annual Proportion AFDC/TANF recipients in County of Residence in GAIN/WTW program	0.309	0.225	0.036	1
Child Care Expend.	Per capita public Expenditures (in 1000s of $1992/93) on Child Care for Low-Income Population in County of Residence county in given year	0.017	0.028	0	0.174
Entered AFDC/TANF in 1986	Household first went on AFDC/TANF in 1986 (= 1, yes; = 0, no)	0.054	0.226	0	1
Entered AFDC/TANF in 1987	Household first went on AFDC/TANF in 1987 (= 1, yes; = 0, no)	0.015	0.121	0	1
Entered AFDC/TANF in 1988	Household first went on AFDC/TANF in 1988 (= 1, yes; = 0, no)	0.023	0.150	0	1
Entered AFDC/TANF in 1989	Household first went on AFDC/TANF in 1989 (= 1, yes; = 0, no)	0.034	0.181	0	1
Entered AFDC/TANF in 1990	Household first went on AFDC/TANF in 1990 (= 1, yes; = 0, no)	0.056	0.230	0	1
Entered AFDC/TANF in 1991	Household first went on AFDC/TANF in 1991 (= 1, yes; = 0, no)	0.090	0.286	0	1
Entered AFDC/TANF in 1992	Household first went on AFDC/TANF in 1992 (= 1, yes; = 0, no)	0.096	0.295	0	1
Entered AFDC/TANF in 1993	Household first went on AFDC/TANF in 1993 (= 1, yes; = 0, no)	0.121	0.326	0	1
Entered AFDC/TANF in 1994	Household first went on AFDC/TANF in 1994 (= 1, yes; = 0, no)	0.117	0.322	0	1
Entered AFDC/TANF in 1995	Household first went on AFDC/TANF in 1995 (= 1, yes; = 0, no)	0.105	0.307	0	1
Entered AFDC/TANF in 1996	Household first went on AFDC/TANF in 1996 (= 1, yes; = 0, no)	0.090	0.286	0	1
Entered AFDC/TANF in 1997	Household first went on AFDC/TANF in 1997 (= 1, yes; = 0, no)	0.078	0.268	0	1
Entered AFDC/TANF in 1998	Household first went on AFDC/TANF in 1998 (= 1, yes; = 0, no)	0.063	0.244	0	1
Entered AFDC/TANF in 1999	Household first went on AFDC/TANF in 1999 (= 1, yes; = 0, no)	0.058	0.233	0	1
Local Labor Market Characteristics:					
Emp-to-Pop	Annual Employment-to-Population ratio in County of Residence	0.617	0.106	0.317	1.330
Prop. Employ. Manuf.	Proportion of County of Residence Employment in Manufacturing in given year	0.135	0.051	0	0.305
Prop. Employ. Services	Proportion of County of Residence Employment in Services in given year	0.292	0.071	0	0.931
Prop. Employ. Retail Trade	Proportion of County of Residence Employment in Retail Trade in given year	0.176	0.026	0.025	0.399
Ave. Earnings, All Sectors	Average Annual Earnings per Worker in County of Residence in 1000s of 1992/93$, All Sectors	34.109	6.297	14.278	78.381
Ave. Earnings, Manuf.	Average Annual Earnings per Worker in County of Residence in 1000s of 1992/93$, Manufacturing Sector	40.993	8.416	0	136.017
Ave. Earnings, Services	Average Annual Earnings per Worker in County of Residence in 1000s of 1992/93$, Service Sector	33.000	6.788	0	81.233
Ave. Earnings, Retail Trade	Average Annual Earnings per Worker in County of Residence in 1000s of 1992/93$, Retail Trade Sector	19.521	2.225	8.705	33.310
Ave. Adult Age	Mean Age of Adults in Household in given year	30.687	8.430	19	103
Female Head	Households is Headed by Female (= 1, yes; = 0, no)	0.928	0.259	0	1
Hispanic	All adults in Household are Hispanic	0.296	0.456	0	1
Black	All adults in Household are black	0.220	0.414	0	1
Asian	All adults in Households are Asian	0.072	0.259	0	1
Other/Mixed Race/Ethnicity	Adults in Household are of other and/or mixed races	0.012	0.111	0	1
County of Residence Dummy Variables:					
Alameda	Resided in Alameda County at sampling	0.043	0.202	0	1
Alpine	Resided in Alpine County at sampling	0.000	0.009	0	1
Amador	Resided in Amador County at sampling	0.001	0.029	0	1
Butte	Resided in Butte County at sampling	0.009	0.094	0	1
Calaveras	Resided in Calaveras County at sampling	0.002	0.039	0	1
Colusa	Resided in Colusa County at sampling	0.001	0.024	0	1

Variable		Mean	Std Dev	Min	Max
Contra Costa	Resided in Contra Costa County at sampling	0.022	0.147	0	1
Del Norte	Resided in Del Norte County at sampling	0.002	0.041	0	1
El Dorado	Resided in El Dorado County at sampling	0.004	0.061	0	1
Fresno	Resided in Fresno County at sampling	0.033	0.179	0	1
Glenn	Resided in Glenn County at sampling	0.001	0.032	0	1
Humboldt	Resided in Humboldt County at sampling	0.006	0.075	0	1
Imperial	Resided in Imperial County at sampling	0.007	0.081	0	1
Inyo	Resided in Inyo County at sampling	0.001	0.027	0	1
Kern	Resided in Kern County at sampling	0.029	0.167	0	1
Kings	Resided in Kings County at sampling	0.006	0.079	0	1
Lake	Resided in Lake County at sampling	0.004	0.061	0	1
Lassen	Resided in Lassen County at sampling	0.001	0.038	0	1
Los Angeles	Resided in Los Angeles County at sampling	0.288	0.453	0	1
Madera	Resided in Madera County at sampling	0.004	0.067	0	1
Marin	Resided in Marin County at sampling	0.002	0.049	0	1
Mariposa	Resided in Mariposa County at sampling	0.001	0.027	0	1
Mendocino	Resided in Mendocino County at sampling	0.004	0.064	0	1
Merced	Resided in Merced County at sampling	0.010	0.101	0	1
Modoc	Resided in Modoc County at sampling	0.001	0.024	0	1
Mono	Resided in Mono County at sampling	0.000	0.014	0	1
Monterey	Resided in Monterey County at sampling	0.013	0.112	0	1
Napa	Resided in Napa County at sampling	0.002	0.049	0	1
Nevada	Resided in Nevada County at sampling	0.002	0.048	0	1
Orange	Resided in Orange County at sampling	0.042	0.200	0	1
Placer	Resided in Placer County at sampling	0.006	0.075	0	1
Plumas	Resided in Plumas County at sampling	0.001	0.029	0	1
Riverside	Resided in Riverside County at sampling	0.051	0.220	0	1
Sacramento	Resided in Sacramento County at sampling	0.052	0.222	0	1
San Benito	Resided in San Benito County at sampling	0.001	0.037	0	1
San Bernardino	Resided in San Bernardino County at sampling	0.083	0.276	0	1
San Diego	Resided in San Diego County at sampling	0.072	0.259	0	1
San Francisco	Resided in San Francisco County at sampling	0.013	0.114	0	1
San Joaquin	Resided in San Joaquin County at sampling	0.024	0.153	0	1
San Luis Obispo	Resided in San Luis Obispo County at sampling	0.005	0.073	0	1
San Mateo	Resided in San Mateo County at sampling	0.008	0.088	0	1
Santa Barbara	Resided in Santa Barbara County at sampling	0.008	0.088	0	1
Santa Clara	Resided in Santa Clara County at sampling	0.032	0.175	0	1
Santa Cruz	Resided in Santa Cruz County at sampling	0.005	0.073	0	1
Shasta	Resided in Shasta County at sampling	0.008	0.091	0	1
Sierra	Resided in Sierra County at sampling	0.000	0.011	0	1
Siskiyou	Resided in Siskiyou County at sampling	0.003	0.054	0	1
Solano	Resided in Solano County at sampling	0.013	0.113	0	1
Sonoma	Resided in Sonoma County at sampling	0.009	0.094	0	1
Stanislaus	Resided in Stanislaus County at sampling	0.018	0.134	0	1
Sutter	Resided in Sutter County at sampling	0.003	0.057	0	1
Tehama	Resided in Tehama County at sampling	0.003	0.053	0	1
Trinity	Resided in Trinity County at sampling	0.001	0.027	0	1
Tulare	Resided in Tulare County at sampling	0.016	0.125	0	1
Tuolumne	Resided in Tuolumne County at sampling	0.002	0.047	0	1
Ventura	Resided in Ventura County at sampling	0.013	0.114	0	1
Yolo	Resided in Yolo County at sampling	0.006	0.074	0	1
Yuba	Resided in Yuba County at sampling	0.004	0.064	0	1

University of California
Los Angeles

Essays on Information, Social Interactions, and the Decisions to Work and Participate in the Earned Income Tax Credit

A dissertation submitted in partial satisfaction
of the requirement for the degree
Doctor of Philosophy in Economics

By

Eduardo Fajnzylber

2004

Chapter 1

Presentation

A permanent dilemma in social policy is how to assist individuals in disadvantaged situations, without creating incentives that tend to perpetuate their economic hardship. The reduction of benefits as the economic situation of a family improves acts as an implicit tax on the efforts to become self-reliant. To avoid this problem, traditional welfare systems have either required recipients to engage in activities that would raise their ability to find employments (in the hopes that better employment opportunities will lead them to leave government assistance for well-paid jobs) or, more recently, have imposed limits in the lifetime period during which a person can receive government aid.[1] One of the main innovations in recent welfare history was the expansion in the early 1990s of a previously small program known as the Earned Income Tax Credit (EITC). This program is focused on families that work but receive a low level of earnings. Given its benefit structure, it provides positive incentives for a person to start working.[2] To apply for the EITC, a family must file an income tax return. As this is not a procedure

[1] Some (like Besley and Coate (1992)) have argued that the level of stigma prevalent in a society can limit the extent to which families without real disadvantages might reduce their work effort in order to qualify for government assistance.

[2] Benefits start at zero for a qualifying family with no earnings, increase proportionally with earnings, reach a plateau and are then slowly phased out with increased earnings. See chapters 2 or 3 for more details on the program.

that low-income families are very familiar with[3], one might expect that it would take some time for families to become aware of the existence of the program, its eligibility requirements and application procedures, to understand its pro-work incentives and to respond to these incentives and find jobs. However, earlier research has found that up to 86% of the eligible population (families with qualifying children with earnings in the appropriate range) was already claiming the credit a few years after its introduction, and significant increases in the labor force participation of single mothers coincided with periods of expansion in the benefits amounts and subsidy rates.

This relatively fast penetration of the program motivates the central questions of the articles in this dissertation: How did individuals learn about the Earned Income Tax Credit, and how this knowledge affected their work decisions? More specifically, in Chapter 2, we will address the question of whether part of the learning process about the EITC and the availability of jobs occurred through the existence of networks at the neighborhood level. If individuals learn from their neighbors, we should observe that the EITC expansions of the early 1990s (which increased the incentives to work) should not only have a direct effect on labor force participation but, as some families start leaving welfare and finding jobs, other families might also increase their work effort. However, as we will discuss in Chapter 2, we will not be able to attribute this indirect effect to an informational story only, given that the potential existence of stigma would generate similar predictions and our data does not allow us to separately identify the two effects.

[3] According to tabulations made by the author based on the CPS, the proportion of families in the 0 to 5% quantile of the earnings distribution who filed a tax return went from 38% in 1992 to 50% in 2001.

In Chapter 3, we explore a second potential explanation for the rapid increase in participation in the EITC, which has a clear interpretation as a formal information transmission channel: the availability of tax preparation services. We will investigate whether families are more likely to participate in the Earned Income Tax Credit when there is a tax preparation service operating in the area.

Together, the two chapters suggest that information is a relevant issue in explaining the relative success of the Earned Income Tax Credit in helping needy families while providing positive incentives to work, and that both formal and informal channels were relevant for the high penetration of the program that we see today.

In chapters 2 and 3, we will discuss the relevant literature to which each essay is contributing. To place the topic of this dissertation in a broader context, we will present in the following section a brief account of the current knowledge in the economic profession about how and when individuals choose to learn from others in their decision-making process. In the final section of this introductory chapter, we will propose future directions for research in the area of information diffusion and learning about government programs.

1.- Information, networks and cascades: A brief summary

The goal of this section is to see what the existing literature on information transmission can inform us about the outcomes from different information diffusion processes concerning program rules, application procedures and job opportunities: what predictions

can be made regarding the equilibrium level of learning in neighbourhoods, about the efficiency of these equilibria, about the determinants of the speed of adjustments, etc.

Economists have become interested in modelling information diffusion from an array of questions: financial markets are often shocked by sudden price changes that cannot be explained by changes in fundamentals; development economists have been interested for a long time in the question of how technological improvements are transmitted across producers, areas or countries; marketing professionals have explored the potential of word-of-mouth advertising to inform potential customers about new products; macroeconomists have looked for informational stories to explain the high volatility of capital flows to emerging markets; labor economists have investigated the role of informal networks as informational channels for job market opportunities, etc. However, little attention has been given to the importance of social communication in the decision to participate in government programs, even if some evidence has pointed out that lack of information or uncertainty about program benefits and rules are often important factors determining participation.[4]

We will first review some of the results of the literature on "informational cascades", which is the expression used for situations in which simple "social learning" rules can lead all individuals to follow the choices made by some early players, even against their private information. As we will see, these phenomena often arise in models where

[4] See chapter 3 for references to the program participation literature.

individuals have to decide among a certain number of known alternatives, but with uncertainty about the individual payoffs of these alternatives.

In the context of the decision to participate in the EITC, one can imagine that even if individuals are aware of the existence of the program, the payoffs might be uncertain before incurring the cost of completing the required forms (the benefits depend on the presence of children, on whether these children qualify as dependents for the credit, on the different sources of family income, and on the EITC table that applies for the particular year and family size). As there are some common features to all families (like the complexity of the forms), there is scope for learning from the choices made by others ("if a family with similar earnings and family structure claimed the credit, then it might be worth doing it myself"). An inefficient informational cascade could occur if early players, based on limited information, choose not to incur the cost of learning about the rules of the EITC, sending the system into an equilibrium where nobody participates. This situation could be reversed with the expansion of benefits (or the opening of tax preparations service in the area) causing some individuals to change their behaviour.

Regarding the question of how individuals find jobs, it is not clear that the previous family of models is adequate, especially since this is a two-sided process: individuals might not know about the existence (or the working conditions) of all job openings and firms face uncertainty about the productivity of job applicants. We will discuss some of

the papers that attempt to model the effect of networks of contacts on job market outcomes.

Informational cascades and heard behavior[5]

The classics articles by Banerjee (1992) and Bikhchandani, Hirshleifer and Welch (1992) introduced the notion of "informational cascades".

Such phenomena usually arise in models where a sequence of agents faces a similar decision between different (discrete) alternatives. The return to the different choices is uncertain in some dimension. As long as there is some component that is common to the payoffs of different individuals (like a common shock to the relative value of the different alternatives) there will be something to be learned from the choices made by others. The basic result is that under some circumstances (for instance when a certain number of initial players, based on their limited information, make the same decision) inference about the underlying payoffs cannot be reversed by a person's own information and therefore, from that moment on, all individuals will follow the recommended choice, creating a (potentially inefficient) informational cascade. Essential to the result is the idea that individuals can observe the choices made by others but not their private signal. This property will fail in general when the agents can somehow infer the personal signals from the actions taken (for instance, when the choice is continuous).[6] When private signals are

[5] See Bikhchandani, Hirshleifer and Welch (1998) for an extended survey and discussion of the literature.

[6] An area where considerable attention has been given to the importance of heard behavior is the study of financial crises. The idea is that institutions considering investing in risky projects might follow the actions taken by early movers, disregarding their own signal about the convenience of the project. Even though some papers (Lee (1993), Avery and Zemsky (1998) and Glosten and Milgrom (1985)) argue that assuming

observed, the relevant information will eventually be learned from past history and the optimal choices will be made in equilibrium. In terms of the signal, the only requirement for the formation of cascades is that they cannot be perfectly conclusive of the optimal choice.

The onset of informational cascades can be exacerbated when individuals differ in the precision of their information and the order of play is endogenous: when individuals can either make a choice or delay their decision (at a cost), some equilibria will be such that less informed individuals will delay until someone else (presumably the better informed agents) make a first move, creating an explosive cascade.

Finally, Bikhchandani, Hirshleifer and Welch (1998) point out that the introduction of costly gathering of private signals can exacerbate the cascading problem, while public or private institutions or technologies that increase the level of information could help prevent the onset of inefficient cascades.

These last set of results are particularly interesting in the context of participation in the EITC, where understanding the rules of the program and learning about how to file a tax return can create high information gathering costs. In that sense, it is quite possible that individuals will rely heavily on the information they can infer from other people's choices, potentiality creating low-participation cascades. The introduction of tax preparation services will greatly reduce this information gathering cost, potentially moving the system away from the inefficient cascade.

---

either continuous investment or the existence of market priced assets, heard behavior does not appear in equilibrium, Chari and Kehoe (2003) show that informational cascades can reappear once the timing of investors is allowed to be endogenously determined.

It is interesting to note that, mirroring part of the econometric discussion on the identification problems of social interaction effects (see Chapter 2), Bikhchandani, Hirshleifer and Welch (1998) note that the phenomena of "convergent behavior" of individuals (or "hearding") can have other explanations outside the informational argument: "individuals might face similar decision problems" (equivalent to the correlated unobservables problem generated by endogenous group formation), "positive payoff externalities", "preference interactions" (like welfare stigma), and "sanctions upon deviants" (equivalent to the unobserved group component problem, which causes a correlation in the choices taken by individuals belonging to the same group).

Social networks and the labor market

Sociologists have long recognized the importance of social networks in the labor market process, documenting that "approximately 50% of all workers currently employed found their jobs through friends and relatives".[7]

In a commonly cited formalization of this argument, Montgomery (1991) models simultaneously the choice of job searching mechanism (formally through advertising or agencies or informally through job referrals by friends or relatives) by employers and job seekers, to understand the relationship between social structure, wages and profits. The two-period model assumes that firms are unable to distinguish between high or low productivity workers before making a wage offer, and that workers from the second

[7] Montgomery (1991).

period may have ties with period-one workers (which can refer them, if requested by their employer). Finally, the productivity level is correlated among tied workers (high productivity workers are more likely to be tied to other high productivity individuals). The results suggest that better connected workers have better outcomes, that firms using informal channels make higher profits and that higher social stratification by ability increases wage dispersion.

More recently, Calvó-Armengol and Jackson (2003) develop a discrete time model that allows for unemployment spells.[8] Unemployed individuals can either receive a job offer with a certain probability or be referred by an employed acquaintance who receives an additional job offer (who randomly picks one person among her available contacts). Finally, employed individuals have an exogenous probability of losing their job each period. The model predicts a positive correlation between agents and across time, as well as a positive correlation between the duration of unemployment and the likelihood of finding a job. A variation of the model allows for endogenous dropping of the labor market network if the expected value of staying on it is exceeded by the cost (education, skills maintenance, opportunity cost, etc.), resulting in higher drop-out rates for initially disadvantaged groups.

An interesting result from these models is the idea that more connected individuals are more likely to find jobs. It would be interesting to extend these models to allow for

---

[8] Other papers that formalize the concept of networks and labor market outcomes in the economics literature are Boorman (1975), Calvó-Armengol and Jackson (2000), Arrow and Borzekowski (2001) and Topa (2001).

voluntary unemployment (individuals might choose to depend on government assistance if the job offers are not a good match for them) and see how exogenous shocks (like the increased incentive to work brought by the expansion of the EITC) will tend to propagate differently across different networks. In particular, one could define networks within neighbourhoods along racial or language lines, and make predictions about the speed of propagation as a function of the "intensity" of the network (concentration of families of the same type in the area).

2.- Directions for future research

Following the suggestions from the information diffusion literature, a natural extension of the results of this dissertation will be to combine the two channels explored in this article (the informal diffusion over networks and the formal effect of tax preparation services) in a unified framework. One can define networks along geographical and racial dimensions, measure the degree of concentration of these networks and see how the system responds to two different classes of shocks: the expansion of the EITC which will tend to lower the reservation wage of informed individuals and the introduction of tax preparation services which will reduce the cost of gathering information about the program, therefore increasing the pro-work incentive effect of the program.

One could analyze the conditions under which the initial system might fall in low-participation cascades and see how these situations might be reversed with the introduction of the incentive and information shocks. Following the idea of Calvó-Armengol and Jackson (2003), it is possible that this type of models might help to explain

the significant participation differences observed between racial groups; it is possible that the structure of the networks formed by minorities will tend to perpetuate low-level equilibria.

As mentioned earlier, it is extremely difficult to distinguish between alternative explanations of the phenomena of "convergent behavior" (why people belonging to the same group tend to make similar choices). While the empirical strategy used in Chapter 2 allows us to eliminate most of the identification problems, we are still left with an interpretation problem: is the positive social interaction observed in the work decisions of welfare families due to informational diffusion or to some form of preference interaction or stigma (individuals might be psychologically affected by departures from the prevalent behavior in the network). In general, these two stories give very similar predictions in terms of the direction of indirect effects: as some individuals react to the EITC expansion by joining the labor market, other individuals might either learn from their predecessors or be compelled for stigmatic reasons to follow their example.

It is possible that by combining social networks with information gathering technologies, we might be able to generate testable predictions that would allow us to give a more precise interpretation to the social interactions effects found in this dissertation.

Chapter 2

The Social Impact of the Earned Income Tax Credit

1.- Introduction

One of the main changes in social policy in recent years has been the expansion of the Earned Income Tax Credit (EITC) in the early 1990s. The EITC is a pro-work social program for families with children. It provides zero benefits to non-working families and as the family enters the labor market, the benefits increase with earnings up to a certain point.[9] Several studies have documented that the EITC could be partly responsible for the substantial increase in the labor force participation of single mothers during the 1990s.[10] The apparent success of the Federal EITC has led several States to implement their own versions of the program.[11] Even though the employment effects seem attractive, the program costs in terms of foregone tax revenues are an important consideration.[12]

[9] See Appendix 1 for a brief description of the EITC. Hotz and Scholz (2000) provide an extensive survey of the research addressing the EITC and its impact on economic and behavioral outcomes.

[10] See for example Eissa and Liebman (1996), Dickert, Houser, and Scholz (1995), Meyer and Rosenbaum (1999), Ellwood (2000), Grogger (2003a) , Grogger (2003b), and Hotz, Mullin, and Scholz (2001).

[11] By 2000, 15 states had implemented some form of State EITC. The most common variant is simply to allow a family to claim a fixed percentage (in general, 10%) of the Federal EITC.

[12] In 1999, for instance, the total Federal expenditure on the EITC was $31.9 billion.

Previous studies have used individual level data to estimate the differential impact of the EITC on families with different characteristics (in this case, family size). We refer to these effects as the *private effects* of the EITC in the sense that they can be attributed to the changes in the incentives faced by each individual family, regardless of the decisions taken by other individuals. However, these studies have not taken into account the indirect effects of the program: encouraging one woman to work may have positive spillovers on other women in the same neighborhood. We refer to these indirect effects as *spillover effects*. These effects, if positive and significant, can generate a *social multiplier* that would amplify the private effect of the EITC.[13] We refer to the overall effect (the sum of the private effect and the spillover effect) as the *social impact* of the EITC.[14] The magnitude of this social multiplier depends on the intensity of the *social interactions* between individuals, i.e. to what extent the fraction of working families in a neighborhood affect an individual's labor force participation decision.

In this article, we first estimate the magnitude of the spillover effect of the EITC, by comparing the labor force participation decision of single mothers living in neighborhoods with different proportions of participating individuals between 1993 and 1994, the period of the largest EITC expansion. We use this estimate to construct a measure of the social impact of the program.

[13] See Glaeser, Sacerdote, and Scheinkman (2002).

[14] The concepts of private and social are relative to the type of variation used to identify the corresponding effects. If for example, the EITC only affected some individuals within a family but not all of them, using family level data would capture the "family social effect" in the sense that it would embed the individual level effects and any type of spillover within the family.

Traditional economic models have assumed that decision making is based on preferences over different goods and constraints imposed by the resources available to an individual (or a family). The total demand for a product is obtained by simple aggregation of individual choices. However, other social sciences, particularly sociology or anthropology, have long believed in the importance of the social environment on the decisions made by individuals. By measuring separately the private and spillover effects of changes in the incentive structure surrounding the EITC expansion, we directly test the hypothesis that the environment of an individual, in this case the neighborhood, can have a direct influence on her decisions.[15]

From a policy perspective, and given the high cost of the EITC, it is important to quantify both the private and the spillover effects of the program. The private effect represents the result of increasing the incentives to work by subsidizing the initial wage perceived by a low income worker, independent of the location of the individual. Predictions of the direct effect of further raising these incentives, for example by introducing a State EITC, could be obtained from these estimates.

---

[15] The experimental evaluation literature on social interactions has found evidence of social effects in retirement plan decisions (Duflo and Saez (2002)), and academic effort and social groups (Sacerdote (2001)). Marmaros and Sacerdote (2002) also find evidence of peer effects in employment outcomes of college graduates from other students in the same dorm hallway. Katz, Kling and Liebman (2000) analyze data from a randomized housing voucher program and find that low income families who moved to low-poverty neighborhoods experienced improvements in different measures of well being but no significant short-run effect on employment, earnings or welfare receipt. Ludwig, Duncan and Pinkston (2000) find some evidence that moving to a low-poverty neighborhood might reduce welfare participation and increase employment.

On the other hand, if the spillover effects are important, it might be a more cost-effective approach to implement policies that act directly on the social interactions. For example a policy designed to increase the flow of information about job opportunities or the benefits of the EITC, especially in low participation neighborhoods, could have large effects on labor force participation, at a relatively low cost.

To estimate the pure effect of interacting individuals, we need to address the identification problems associated with trying to explain individual behavior with mean behavior in the individual's reference group (Manski's "reflection problem"[16]). As the outcome of individual B determines the outcome of individual A in the same neighborhood (and vice versa), we have a standard *simultaneity problem*. Depending on the nature of the social interactions (endogenous or exogenous), the reduced form will not identify all the structural coefficients.[17] The second problem, *endogeneity caused by correlated unobservables*,[18] occurs when individual characteristics (specially the unobserved ones) are correlated among individuals belonging to a certain group. This case is of particular concern when individuals self-select into groups (*endogenous membership*). Even in the absence of social interactions, similar individuals will tend to make similar choices, causing a spurious correlation between individual and group

[16] See Manski (1993).

[17] Manski (1993) calls *endogenous* the effects of average behavior on individuals decisions and *exogenous*, the effect of mean exogenous characteristics of the individuals in a person's reference group.

[18] See Manski (1993) or Moffitt (forthcoming).

behavior. The third problem, *unobserved group characteristics*, originates from the difficulty in controlling for all neighborhood characteristics that affect the outcome of interest. Those omitted characteristics could generate a correlation in the decisions taken by individuals in the same group that is unrelated to any form of social interaction.

Previous studies have dealt in different ways with the reflection and unobserved neighborhood characteristic problems. We discuss some of these approaches in section 2. The correlated unobservables problem has remained a difficult issue to control directly.[19]

We use the EITC expansion between 1993 and 1995 that differentially affected families with one child versus families with two or more children to overcome the unobserved group component and reflection problems. [20] [21] The average fraction of one child versus two-or-more children families in a neighborhood (interacted with a 1994 year dummy variable) can be used as an instrument to isolate the impact of the average labor force participation in the neighborhood on individual behavior. The idea is that a woman living in a neighborhood with a higher fraction of two-or-more-children is more likely to be "exposed" to a higher increase in the average labor force participation rate than a woman

---

[19] An exception is Moretti (forthcoming), who includes individual*city fixed effect to account for sorting of high ability workers into cities.

[20] This idea was originally mentioned in Moffitt (forthcoming), who suggested that "...partial-population experiments in which only a portion of the individuals within each group are given a treatment..."\ could provide an identification strategy that didn't involve experimental manipulation of group membership (Page 20).

[21] The maximum benefit for families with two children rose from $1511 in 1993 to $3110 in 1995, while families with one child increased their maximum benefit from $1434 to $2094 in the same period.

living in a neighborhood with fewer two-or-more-children families. However, the EITC expansion is not a perfect experiment. The intensity of the treatment is a function of family size, which might be correlated with unobserved determinants of labor force participation. Self selection into neighborhoods potentially will cause these unobserved factors to correlate among individuals. We need a strategy to deal with the endogenous membership problem.

Given that endogenous membership is essentially a sample selection problem, we use an individual fixed effect approach on a two-period (1993 and 1994) longitudinal data set composed of single mothers participating on welfare at some point between 1992 and 1995.[22] Assuming that, in the short run, the distribution of individuals within neighborhoods remains stationary, the correlated unobservables problem caused by endogenous membership is "differenced out" by the individual fixed effect for those individuals that did not change their zip code. Assuming that family size is uncorrelated with unobserved determinants of employment (conditional on the individual fixed effect), the average fraction of one versus two-or-more children families in a neighborhood will be a valid instrument for the changes in the average labor force participation. Additional controls for local economic conditions (at the county level), will eliminate most of the remaining endogeneity problems caused by unobserved neighborhood characteristics. We also estimate models that more robustly control for changes in local economic conditions

[22] This strategy is proposed in Kyriazidou (1997). Brock and Durlauf (2001) suggest that panel data, under appropriate conditions, could help solve the self-selection problem in interactions-based environments (page 3337).

or changes in the work emphasis of welfare offices of different counties, by including county-time fixed effects.

We use administrative data from the state of California, which provides the universe of women receiving welfare benefits in California, at some point between 1992 and 1995. Our neighborhood definition corresponds to the zip code of residence of these women and their families.

We estimate a private effect of the EITC of the same magnitude as found in previous studies. Between 1993 and 1994, single mothers with two or more children experienced an increase in their labor force participation by approximately 1.6 percentage points more than families with only one child. As we show, this estimate is robust to the inclusion of the social interaction effect and to different sample definitions. The aggregate private impact of the program was to raise the overall labor force participation among low income mothers in California by 1.1 percentage points.

The estimated magnitude of the spillover effect (pure social interaction) ranges from 0.46 to 0.63, which implies that if a woman was living in a neighborhood that increased its average labor force participation from 50% to 60% between 1993 and 1994, her likelihood of working would increase by 4.6 to 6.3 percentage points. Our combined effect, the aggregate social impact of the EITC between 1993 and 1994, ranges from 2.2 to 3.2 percentage points. The spillover effects are, however, less precisely estimated.

The remainder of the paper is structured as follows: in section 2, we present the basic specification that we will estimate in this paper, laying out the identification problems and the strategy we will follow to overcome them. In section 3 we present the data. In section 4, we present and discuss our results and some robustness analyses and in section 5 we conclude.

Appendix A presents a brief summary of the main characteristics of the EITC program, Appendix B discusses the assumptions for identification of the linear social interactions model with individual fixed effects and instrumental variables and in Appendix C, we present a simple model of location and work decisions that illustrates how the endogenous sorting into neighborhoods can generate a correlation between individual and neighborhood averages, even in the absence of social interaction effects.

2.- Empirical Identification Strategy

As a benchmark, we will estimate the private effect of the EITC using a similar model to the one used in the evaluation literature on the EITC, which exploits the differential treatment of the program for families with one child and families with two children and how this differential treatment was expanded between 1993 and 1994. This essentially corresponds to a difference in difference model including individual level characteristics as controls. The model is estimated for a sample of single mothers with children.

$$
\begin{aligned}
&LFP_{igt} = \beta + \gamma\left(KIDS_{it}^{2+} * YEAR_{it}^{1994}\right) + \delta_1 KIDS_{it}^{2+} + \delta_2 YEAR_{it}^{1994} \\
&+\delta_3 AGE_{it} + \theta' LEC_{gt} + \varepsilon_{it} \\
&t = 1993,1994 \\
&i = 1..N
\end{aligned} \tag{1.1}
$$

where LFP_{igt} corresponds to the labor force participation of woman *i* who lives in neighborhood *g* in time *t*, $YEAR_{it}^{1994}$ is a dummy variable equal to 1 if the observation corresponds to 1994 and zero otherwise, $KIDS_{it}^{2+}$ is a dummy variable equal to 1 if the woman has 2 or more children and 0 if the woman has only one child, AGE_{it} is the age of the mother at the end of year *t*, AGE_{it}^2 is the squared age of the mother, and LEC_{gt} is a measure of Local Economic Conditions in the area (county) where individual *i* resides in period *t*.

Under reasonable conditions, the ols estimate γ will consistently estimate the private effect of the EITC.[23]

To simplify notation, we will rewrite equation (1.1) as:

$$Y_{igt} = \alpha + \gamma EITC_{it} + \delta' X_{it} + \theta' Z_{gt} + \varepsilon_{igt}$$

[23] See Hotz and Scholz (2000) for a discussion of the conditions for identification in this model.

where Y_{igt} is the labor force participation decision, $EITC_{it}$ corresponds to the interaction term, X_{it} is a vector of remaining individual characteristics and Z_{gt} represents neighborhood characteristics (observed).

Now, let's introduce the spillover effect of the EITC. The simplest way to model such effects is to include the average participation among individuals living in the same area as a regressor.[24]

$$Y_{igt} = \alpha + \beta E\left[Y_{it} \mid G_{it}\right] + \gamma EITC_{it} + \delta' X_{it} + \theta' Z_{gt} + \left(\mu_{gt} + E\left[\varepsilon_{igt} \mid G_{it}\right] + \nu_{igt}\right) \quad (1.2)$$

where G_{it} represents the neighborhood where individual *i* resides in period *t* and $E\left[Y_{it} \mid G_{it}\right]$ corresponds to the population mean labor force participation in neighborhood G_{it}.

Notice that we have assumed that the average EITC variable in the neighborhood does not directly affect the individual decisions. In this case, this is a reasonable assumption to the extent that the EITC expansion was an unanticipated event and that in the short run, we assume that neighborhoods will not change in response to the policy.

We decomposed the error term in three terms, to reflect the main identification issues associated with the estimation of social interactions models:

[24] This assumption, the most common in the empirical literature, can be justified from the linearized version of a model of information diffusion (also known as contagion models). The probability of meeting an informed individual is a function of the average participation in a neighborhood.

- The term μ_{gt} represents *unobserved neighborhood characteristics*. As these components will affect each individual in the neighborhood in a similar way, it will necessarily be related to the average participation in the neighborhood.

- The second term, $E\left[\varepsilon_{igt} \mid G_{it}\right]$, represents the problem of *correlated unobservables*; Individuals who **chose** to live in the same neighborhood (who have the same G_{it}) will potentially share common unobserved determinants of labor force participation. If not accounted, the existence of this term will generate a correlation in the decisions of individuals, even in the absence of social interactions. In appendix C, we present a simple location/work model to illustrate this issue.

- Finally, ν_{igt} is an error term, assumed uncorrelated with the regressors in the equation and the determinants of the location decision.

In practice, $E\left[\varepsilon_{igt} \mid G_{it}\right]$ is replaced by the sample mean $\overline{Y_{G_{it}t}} = \frac{1}{N_{G_{it}}} \sum_{G_{jt}=G_{it}} Y_{jt}$. This will generate a standard *simultaneity problem*: If individuals A and B live in the same neighborhood, their LFP decisions will enter each other's equation.

In order to deal with the simultaneity and the unobserved group component problems, some studies have used different measures of the network available to an individual in a particular neighborhood:

- Bertrand et al. (2002) use an interaction between "contact availability" (the proportion of individuals that speak the same language in the neighborhood) and "quality of contacts" (average welfare use of the language group in the country):

$$Y_{ijg} = \beta\left(CA_{jg} * \overline{Y_j}\right) + \delta' X_i + \gamma_g + \eta_j + \theta' CA_{jg} + \nu_{igt}$$

where *j* represents the language group and *g* the neighborhood. As the interaction term varies for individuals from different language groups in the same neighborhood, they are able to include a neighborhood fixed effect (γ_g) that will capture any unobserved neighborhood characteristic (that is not group specific).

- Aizer and Currie (2002) estimate models where networks are measured as the average outcome (utilization of publicly funded prenatal care) among women in the same neighborhood and race group during the 11 months prior to a birth:

$$Y_{ijgt} = \beta_1\left(CA_{jgt} * \overline{Y_{jg(t-1)}}\right) + \beta_2 \overline{Y_{jg(t-1)}} + \beta_3 \overline{Y_{(-j)g(t-1)}} + \delta' X_i + \gamma_{gt} + \nu_{igt}$$

where *j* and *-j* represent the "own" and "others" language groups, *g* represents the neighborhood and γ_{gt} is a neighborhood-time fixed effect that will capture time-varying unobserved group characteristics. This dynamic model, with lagged outcomes on the RHS of the equation, is no longer affected by the simultaneity problem.

To identify the coefficients of interest (β and γ), we will follow a different approach. We first obtain a reduced form equation by taking conditional expectations over both sides of equation (1.2) and solving for $E\left[Y_{it} \mid G_{it}\right]$:

$$E\left[Y_{it} \mid G_{it}\right] = \frac{\alpha}{1-\beta} + \frac{\gamma E\left[EITC_{it} \mid G_{it}\right]}{1-\beta} + \frac{\delta' E\left[X_{it} \mid G_{it}\right]}{1-\beta} + \frac{\theta' Z_{gt}}{1-\beta} + \frac{\mu_{gt} + E\left[\varepsilon_{igt} \mid G_{it}\right]}{1-\beta} \quad (1.3)$$

This represents the unique Social Interaction equilibrium of the system. Estimating this equation, it is possible to obtain the Social impact of the EITC as the coefficient corresponding to $E\left[EITC_{it} \mid G_{it}\right]$.

$$\text{Social Impact of the EITC} = \frac{\gamma}{1-\beta} = \gamma + \frac{\beta\gamma}{1-\beta} \quad (1.4)$$

To decompose the effect into a private (γ) and a spillover effect ($\frac{\beta\gamma}{1-\beta}$), we need to go back to the individual level equation, where we can $E\left[Y_{it} \mid G_{it}\right]$ from equation (1.3).

$$\begin{aligned} Y_{it} &= \frac{\alpha}{1-\beta} + \frac{\gamma\beta}{1-\beta} E\left[EITC_{it} \mid G_{it}\right] \\ &+ \frac{\beta\delta' E\left[X_{it} \mid G_{it}\right]}{1-\beta} + \gamma EITC_{it} + \delta' X_{it} + \frac{\theta' Z_{gt}}{1-\beta} + \left(\frac{\mu_{gt} + E\left[\varepsilon_{igt} \mid G_{it}\right]}{1-\beta} + \nu_{igt} \right) \end{aligned} \quad (1.5)$$

As we discuss in appendix B, assuming exogeneity of the observed characteristics (including $EITC_{it}$) and exogenous allocation into neighborhoods, it is possible to identify

all the coefficients from this reduced form equation (1.5). Imposing the restrictions of the Social equilibrium (1.3) could even lead to overidentification of the model. It is, however, difficult to justify the exogeneity assumptions. The individual observed characteristics (X_{it}) are likely to be correlated with determinants of location decisions or with unobserved neighborhood characteristics. Even the EITC variable (which is a function of family size) could be correlated with unobserved determinants of the location decision.

For these reasons, we will exploit the panel nature of our data set by including an individual fixed effect and restricting our analysis to the subsample of individuals that didn't change neighborhoods. Instead of estimating the two reduced form equations, we will use a two stage least squares estimation procedure where the equilibrium equation will be estimated in the first stage and the predicted average LFP will replace $E\left[Y_{it} \mid G_{it}\right]$ in equation (1.2). As we use a two-year panel of individuals (and we focus on the individuals who didn't move between 1993 and 1994), we can express the estimating model in first differences, in the following system:

$$\text{First Stage} \quad : \Delta\overline{Y_{gt}} = \pi_0 + \pi_1 \Delta\overline{EITC_{gt}} + \pi_2 '\Delta X_{it} + \pi_3 '\Delta Z_{gt} + \eta_{gt}$$

$$\text{Second Stage} : \Delta Y_{igt} = \alpha_0 + \beta\widehat{\Delta\overline{Y_{gt}}} + \gamma EITC_{it} + \delta '\Delta X_{it} + \theta '\Delta Z_{gt} + \Delta\varepsilon_{igt}$$

This procedure, under the assumptions discussed in Appendix B, will serve three purposes:

- Capture time invariant neighborhood characteristics,

- Turn the EITC variable (conditional on the individual fixed effect) into a valid instrument for time varying unobserved group characteristics,

- Eliminate the selectivity bias introduced by endogenous sorting into neighborhoods.

As we will see in section 4.4, we also present some estimates based on the individual level reduced form equations, assuming that the system is in a social interaction equilibrium. In section 4.5, we repeat both estimation strategies for a model with county-time fixed effects, which allows us to control more effectively for determinants of work that might be changing over time at the county level, including local economic conditions and the particular emphasis on work requirements of the different counties.

3.- The Data

Estimation of the proposed identification strategy required availability of a very particular data set which would contain geographical information (zip code) on a high number of individuals belonging to an homogeneous population, with at least two years of longitudinal data.

The data set used is based on confidential administrative data from the State of California. The main source corresponds to a series of yearly files from the MediCal Eligibility Data System (MEDS), which keeps monthly eligibility codes for every individual eligible for MediCal, the State health insurance program. As individuals participating on the AFDC program are automatically incorporated into the MEDS file, this could be considered as a census of all individuals receiving welfare assistance in California. We had access to these files for the period 1987-2001 but in order to justify some of the assumptions required for identification (that neighborhoods didn't change dramatically in response to the EITC expansion), the analysis will be limited to the 1993-1994 period, where the biggest expansions of the EITC benefits took place.

As we was interested in a sample of single mothers, we constructed a sample of women with ages between 19 and 45, in "cases" where no other person over 18 is present, where at least one person under 18 is present and the oldest person under 18 has a difference of at least 13 years with the adult.

The MEDS file provides monthly information on the eligibility codes and county of residence, zip code information every six months, and a limited number of demographic characteristics like date of birth, race and language. One crucial aspect of this data is that family composition and geographical information are only available for those periods where the person was eligible for MediCal. Our base sample corresponds to all the women that were present at least once during the period 1992-1995 and we applied a

simple algorithm to impute location and number of children for the years in which the person was not observed:

- Impute family composition (number and ages of children) and geographic location (zip code) backwards. If, for example, the person was on MediCal during 1993 and 1995 but not in 1994, we impute the first zip code observed in 1995 as the location for 1994 and we roll back the ages of children present during the first month of 1995 to obtain imputed family structure for 1994.

- If no information is available from the future (up to 1995), then we impute forward. If a woman was on MediCal during 1994 but not in 1995, we will impute the zip code in 1995 from the last zip code observation in 1994. We will also roll forward the ages of children from the last observation in 1994.

In order to construct labor force participation for these individuals, we used matched information from the California Unemployment Insurance administration, which corresponds to quarterly data on the earnings received by each person from any job covered by this program. This only excludes government jobs and self-employment.[25] To avoid problems with incorrect social security numbers (that could result in non matches) or numbers that were being used by multiple individuals, we restricted the sample to

[25] In the future, this data will also be merged with tax information from the California Franchise Tax Board, which will allow us to include earnings from other sources like self-employment. However, there is some

individuals with verified identifiers and with less than 12 employers in a year. Finally, we discarded families with more than 6 children.

With this sample, we constructed our basic measure of individual labor force participation, which is a dummy variable equal to 1 if annual earnings were higher than $200.

Notice that labor force participation is never imputed because we observe earnings independently of MediCal eligibility.

We also used county information to construct measures of local economic conditions: aggregate data from the California Unemployment Insurance Administration provides yearly measures of employment (number of individuals) by county at 10 different industrial sectors. We used both the overall employment by county and employment in the retail and services sectors, which are traditional sources of jobs for the low-income population. Combined with information from the California Department of Finance for the population aged 18-65 by county, we constructed employment to population ratios for the three sectors mentioned above (all sectors, retail and services). We also constructed measures of average earnings for all sectors and for the retail and service sectors. This measure is constructed by dividing a measure of payroll by a measure of employment.

---

evidence that this population is particularly prone to not file income tax returns, either because of eligibility requirements, lack of information or fear of being audited.

As an alternative measure, we also matched unemployment rates at the county level from the Bureau of Labor Statistics.

3.1.- Summary Statistics

In table 2, we present some summary statistics of the individuals in our baseline sample. Each observation corresponds to a single mother with at least one child. Besides from the validation conditions mentioned in the previous section, the baseline sample corresponds to all those individuals that didn't change zip code between 1993 and 1994, who lived in zip codes with at least 101 individuals (in our sample) and that didn't change the number of children between 1993 and 1994. The restriction to a population of "stayers" is motivated by our identification strategy; for this sample, the individual fixed effect will effectively "difference out" the selectivity bias. The minimum sample size restriction was introduced to minimize the measurement error inherent in estimating population means with sample means. Finally, the restriction in fertility is a way of controlling for changes in the EITC eligibility group; a woman who goes from having one child (low benefit group for the EITC) to having two children (high EITC eligibility group) is much more likely to reduce her labor force participation during the first year of life of the newborn.

Our data set includes 142391 individuals distributed in 503 zip codes. Average labor force participation rose between 1993 and 1994 from 33.7% to 36.7%. Approximately 69.1 % of these families have more than 1 child, with an average of 2.2 children. The

sample is also predominantly Hispanic (42.8%) and white (28.8%) and the average age of mothers is almost 34. As can be seen from the measures of local economic conditions, the economy was mostly improving in California with both employment increasing and unemployment falling.

We can also see that both the fraction of women working in a neighborhood and the fraction of one v/s two-or-more children families have a reasonable range of variation: average LFP goes from 13.8% in some areas to 58% in others, while the fraction of 2-or-more-children families varies from 0.386 to 0.838.

In terms of our estimation strategy, more relevant than the average LFP in a zip code is the change in that variable between 1993 and 1994. We present the distribution of this change in figure 3 and the distribution of the fraction of 1 vs. 2 or more children families in figure 4. From figure 3, we notice that most neighborhoods raised their average LFP between 0 and 13 percentage points. There is also a certain fraction of areas that decreased their average participation.

4.- Results

4.1.- The Private Effect of the EITC

As a benchmark model, we present in table 3 the results of estimating the private effect of the EITC using our sample of single mothers. The first column corresponds to a cross sectional model for 1994. After controlling for age and age squared, the labor force participation of women with more than one child is 2.7 percent lower than that of women with two children. So, in general, larger families tend to work less than families with only one child.

The second column presents the results for a model where data from 1993 and 1994 were pooled together but instead of including a fixed effect, we allowed for clustering at the individual level. From this model, we obtain our first estimate of the private effect of the EITC, given by the coefficient on the interaction between the “More than 1 child” dummy variable and the “Post 1993” dummy variable. The results suggest that women with 2 or more children raised their labor force participation between 1993 and 1994 by 1.6 % more than women with only one child. This is a very similar result to the ones found in the evaluation literature of the EITC.

The other coefficients have the expected signs: the direct effect of having more children tends to lower participation but this variable tends to increase with age at a decreasing

rate. The coefficient on the employment to population ratio, however, has the wrong sign.[26]

Finally, the third column presents the equivalent to column two with individual fixed effects, with extremely similar results. As we restricted our sample to women that didn't change their number of children, the direct effect of having more than 1 child is completely difference out.

4.2.- Private Effects, Spillovers and the Social Impact of the EITC

Our main results are presented in table 4, which includes the estimates from three different models that allow for social interactions by including the average labor force participation on the RHS of the equation.

The first column corresponds to the estimation of a model that includes a fixed effect (that should eliminate the selectivity bias) but that doesn't include an instrument for the average participation in the neighborhood. As we explained earlier in this article, failing to solve the simultaneity and unobserved group components problems can result in a strong endogeneity problem that will tend to bias upwards the social interaction

[26] We experimented with different controls for local economic conditions, with practically no effect on the coefficients of interest. We only present the results with the most general control, the employment to population ratio for all sectors.

coefficient.[27] In this case, the estimate is practically equal to one and extremely significant. Notice that the estimated private effect (presented in the second row of the table) remained practically unchanged from the results in table 3. This suggests that if the object of interest of the evaluation of a program is to obtain the private effect (rather than the social effect), it sounds reasonable to perform an individual evaluation without having to worry about the spillover effects. Put in other terms, the spillover effect cannot be viewed as an omitted variable problem that will bias the results of a private evaluation.

The second column presents the result of the opposite exercise, in which we use the instrument generated by the expansion of the EITC but we do not include fixed effects.[28] This can be interpreted as a model in which we assume exogenous membership and exogeneity of the included regressors (like in the first case treated in Appendix B). Once again, if individuals who sort into the same neighborhoods tend to have similar propensity to work (or to participate on welfare programs), failure to account for selectivity effects will tend to bias upwards the spillover effects. In this case, we estimate a social interaction coefficient of 0.753, which could be interpreted by saying that if a woman was living in a neighborhood that increased its average labor force participation

[27] In the extreme case of running a regression of an outcome against the average outcome, without regressors, Manski (1993) showed that the ols coefficient would mechanically equal to one, independently of the existence of social interactions.

[28] To make the results more comparable across specifications, we used the exact same sample in each case, which includes the restriction that individuals do not change zip code between 1993 and 1994. This is a restriction imposed in conjunction with the fixed effect approach. The results from the second model are very similar when using the entire sample.

from 50% to 60% between 1993 and 1994, her likelihood of working would increase by 7.53 percentage points.

The other coefficients are in line with previous results (including a private effect of the EITC of 1.5%).

The first results based on our identification strategy are presented in the third column of table 4. In this baseline model, we use both the instrument and the individual fixed effects. Our results suggest a private effect of the EITC of 1.6%, a spillover coefficient of 0.651 and a social impact for the EITC (given by the first stage coefficient presented in the bottom panel of column three) of 0.046. Multiplying this coefficient by the average fraction of 1 vs. 2 or more children families in our sample (0.691), we obtain the overall social impact of the EITC for California, between 1993 and 1994 (standard errors in parenthesis):

$$\text{Overall Social Impact of the EITC} = \hat{\pi}_1 * \Delta\overline{\overline{EITC}} = 0.046 * 0.691 = 0.032\ (0.014)$$

$$\text{Overall Private effect of the EITC} = \hat{\gamma} * \Delta\overline{\overline{EITC}} = 0.016 * 0.691 = 0.011\ (0.002)$$

$$\text{Overall Spillover effect of the EITC} = 0.032 - 0.011 = 0.021\ (0.014)$$

Notice that following the decomposition presented in equation (1.4), we obtain a very similar result for the spillover effect:

$$\text{Overall Spillover effect of the EITC} = \frac{\beta\gamma}{1-\beta}\overline{\overline{\Delta \text{EITC}}} = \frac{0.651*0.016}{1\text{-}0.651} * 0.691 = 0.0206$$

Along with the number of observations and R-squared of both stages, we report both the R-squared associated with the average EITC variable (the excluded instrument). The associated R-squared (with the included regressors partialled out) is 0.01, which is in the range where one would question the strength of the instrument.

4.3.- Sensitivity Analysis

We estimated the baseline specification varying two characteristics of the sample selection procedure: in the first one, we vary the minimum cell size per zip code (the minimum number of individuals that we allow in a particular zip code) and in the second one, we restrict the sample to individuals for which we have multiple observation from which to impute missing location or family composition. The baseline result is reproduced in the third column of table 5.

The first sensitivity analysis is presented in the first three columns of table 5. In the first column, we allow for a minimum cell size of 26 individuals, which more than triplicates the number of zip codes for estimation. We obtain a smaller social interaction coefficient (0.574) and a slightly larger EITC coefficient. Similarly, when we allow for a minimum size of 51 individuals per zip code, we obtain a smaller coefficient than with the baseline sample.

As we mentioned earlier, our base sample include all mothers who were eligible for MediCal at least once between 1992 and 1995 and location and family size is imputed for

the missing years. This imputation algorithm require stronger assumptions in cases where a person is observed only once during the four years. For this reason, we constructed two additional subsamples:

- Sample B: On MediCal during the periods (1993 and 1994) or (1992 and 1994) or (1993 and 1995) or (1992 and 1995).

- Sample C: On MediCal during the periods (1993 and 1994) or (1992 and 1994) or (1993 and 1995).

The idea of these samples is that we only keep individuals for which we have at least two pair of non consecutive observations to impute data for 1993 and 1994.

The results are presented in the last three columns of table 5. The social interaction coefficient have a similar magnitude (slightly smaller) than the baseline results but with larger standard errors. This is quite likely explained by the diminished number of zip codes available to identify the social effect in the first stage (the number of zip codes falls from 503 to 364 when going from the baseline to Sample B and from there to 354 zip codes in sample C).

Overall, our results seem robust to the choice of sample that we used, even though more stringent requirements diminish the effective sample size to levels where our coefficients lose statistical significance.

4.4.- Reduced Form Estimates

The previous estimates were based on the traditional assumptions of instrumental variable estimators for the model in first differences: that the instrument $\Delta\overline{EITC_{gt}}$ is correlated with the endogenous variable $\Delta\overline{Y_{gt}}$ while uncorrelated with the error term in the main equation. The first condition is motivated by the fact that the individual EITC variable determines individual participation, so we would expect that the aggregate EITC should explain average participation in the neighborhood. We did not impose any restriction on the coefficients of either stage.

However, if we assume that the system has reached the social interaction equilibrium of equation (1.3), we can use the Reduced Form equation (1.5) to solve for the social interaction coefficient (β):

$$\Delta Y_{igt} = \pi_0 + \pi_1 \Delta\overline{EITC_{gt}} + \pi_2 \Delta EITC_{it} + \pi_3 '\Delta X_{it} + \pi_4 '\Delta Z_{gt} + \eta_{gt} \qquad (1.6)$$

where

$$\beta = \frac{\pi_1}{\pi_1 + \pi_2}$$

$$\gamma = \pi_2$$

The results for this procedure are presented in table 6, where the upper panel corresponds to the coefficient from the reduced form equation (1.6) for the different samples defined in the previous section, and the lower panel presents the estimates of β using the previous formula. Comparing these estimates with the first row of table 5, we can see that the two procedures give very similar results, with the reduced form coefficients being slightly higher than the 2SLS ones. This suggests that the social interaction equilibrium assumption seems to be a valid one.

4.4.- Models with County-Time Fixed Effects

In the previous specifications, we were controlling for local economic conditions by including a proxy variable for changes in the labor demand at the county level (employment to population ratio). As this measure, or any other, is potentially subject to criticisms, we introduce county-time fixed effect to account for any determinant of labor force participation that may be varying at the county level. This includes, for example, any changes in the emphasis towards work from the county welfare offices.

The results from these specifications are presented in tables 7 (two-stage least squares estimates) and 8 (reduced form estimates).

The results for the social interaction coefficient are smaller than in the previous specifications, suggesting that part of the estimated effect was in fact due to unobserved changes in labor conditions in the local geographic areas. The estimate based on zip codes with at least 100 individuals corresponds to 0.606 (t-stat=3.11) in the model with fixed effects and to 0.651 (t-stat=4.07) in the previous model.

As before, the social interaction coefficient from the sample with at least 50 individuals per zip code is significantly smaller than the baseline sample (with at least 100 individuals per zip code) and is not statistically significant at the 5% confidence level. The implied overall Social, Private and Spillover effects of the program for this sample (two-stage least squares, more than 50 individuals per zip code) are the following (standard errors in parenthesis):

$$\text{Overall Social Impact of the EITC} = 0.032 * 0.691 = 0.022 \; (0.013)$$
$$\text{Overall Private effect of the EITC} = 0.016 * 0.691 = 0.011 \; (0.002)$$
$$\text{Overall Spillover effect of the EITC} = 0.022 - 0.011 = 0.011 \; (0.013)$$

As with previous specifications, the R-squared associated with the excluded regressors are small (0.004 for the case with more than 50 individuals per zip code), suggesting a potentially weak instrument problem. Assuming that the instrument is a valid one (uncorrelated with the main error term), we have to be careful with the statistical inference based on the t-statistic, which might be incorrect for small samples.

5.- Conclusions

In this article, we analyze the social impact of the EITC, which includes both the direct effect that has been repeatedly found in earlier evaluations of the EITC, and the social interaction or spillover effect that results when some women start leaving welfare for work, leading other women to follow their steps. Our direct results are extremely robust and similar to the earlier literature but our social interaction estimates suggest that the indirect effects (multiplier) could double the direct impact, implying that the EITC expansion could be responsible for an overall increase between 2.2 and 3.1 percentage points in the labor force participation of low income single mothers in California. The spillover effects are less precisely estimated, however, being in some cases not significantly different from zero.

An open question in this, as in other social interaction studies, is related to the channel through which these social interactions are operating. The EITC is a very expensive program and it is not clear whether expanding it (for instance through a State EITC) would produce similar results on labor force participation to when the program was initially implemented. If we were able to find that diffusion of information plays an important role, it would be possible to design more cost-effective policies that would operate directly on the social interactions (like job search services in low participation areas). In the research agenda on this topic, we will study whether the existence of tax preparation services (TPS), one of the main avenues through which low income

individuals claim the EITC, has a role as a facilitator of information transmission; the easiest way to convey information about how to apply for the EITC is to refer other people to the closest TPS in the area.

Another implication of these findings is related to the topic of evaluation of government programs. Quite possibly, few things influence more the policy discussion on government programs than experimental evaluations. Given the complex selection issues involved in voluntary participation, a policy experiment has many desirable properties. However, a critical and often overlooked assumption behind social experiments is the non-existence of social interaction effects.[29] Given the magnitude of the spillover effects found in this paper, it is definitely an area to explore more carefully.

[29] See for example Heckman and Smith (1995) or Burtles (1995) for two discussions about the validity of experimental evaluation of social programs.

Appendix A: Overview of the EITC.[30]

The Earned Income Tax Credit is a Federal Program created in 1975, which provides a subsidy for low-income families with children. Initially small, the program was greatly expanded in the early 1990s to become the largest cash transfer program in the US.

It is operated through the income tax system; a family (either a married couple or a head of household) with children that had some work income during a year can file, even if they do not owe any taxes, an income tax form claiming the credit. The benefit works as a refundable credit: if the amount of the benefit is higher than the tax liability, a check is issued for the difference.

In Figure 1, we show the benefit schedule of the program. Starting at zero earnings, the level of benefits increases with earnings up to a certain point. This is called the "phase-in range". Benefits remain constant during the "flat range" and are then reduced during the "phase-out" range.

In contrast with traditional social programs, the EITC provides an unambiguous incentive to participate in the labor market, at least for single headed households. This can be seen in Figure 2, where we show the traditional consumption-leisure trade-off model,

[30] See Hotz and Scholz (2000) for an extensive review of the EITC, its history and behavioral effects.

comparing the regular budget constraint (without EITC) and the modified budget constraint.

Two additional characteristics should be mentioned about the program that separate it from traditional welfare programs: the program does not distinguish between single parents or married couples and the benefits are capped for family size, i.e. the amount is the same for families with 2 or more children.

Appendix B: Identification of Social Interaction effects with Panel Data

In this appendix, we will make a first attempt to establish the conditions under which a fixed effect approach at the individual level can arguably reduce or eliminate the endogeneity bias caused by endogenous membership.[31]

We will first discuss the simple case with exogenous membership, after which we will analyze the more realistic case of endogenous self-selection into neighborhoods.

B.1.- Exogenous Membership and Exogenous Observed Characteristics

The simplest model of social interaction would assume that individuals are exogenously assigned into neighborhoods and that the observed determinants (X) are uncorrelated with the unobserved determinants. By exogenous neighborhoods, we mean that the location process is independent of the observed (X) or unobserved determinants (ε) of the outcome of interest and the unobserved neighborhood characteristics (μ).

$$\begin{aligned} & Y_{igt} = \alpha + \beta \overline{Y_{gt}} + \gamma X_{igt} + \left(\mu_{gt} + \varepsilon_{igt} \right) \\ & E\left[\mu_{gt} \mid \left\{ X_{igt} \right\}_{i=1}^{N_{gt}}, G_{it} = g \right] = 0 \\ & E\left[\varepsilon_{igt} \mid \left\{ X_{igt} \right\}_{i=1}^{N_{gt}}, G_{it} = g \right] = 0 \\ & G_{igt} \perp \varepsilon_{igt}, \mu_{gt}, X_{igt} \end{aligned} \tag{2.1}$$

[31] It is also possible that, under more stringent conditions, a repeated cross-section approach might be able to identify the social interaction effects. We will not discuss that possibility in this article.

where Y_{igt} is the outcome of interest, $\overline{Y_{gt}}$ is the contemporaneous mean of the outcome variable, X_{igt} is an individual determinant of the decision, ε_{igt} is an error term, and μ_{gt} represents unobserved group components. β represents the (endogenous) social interaction effect. G_{it} represents the location of individual *i* in time *t*. N_{gt} represents the number of individuals in neighborhood *g* at time *t*.

Even under this situation, equation (2.1) cannot be consistently estimated by ols, because of the simultaneity of the system of equations for individuals belonging to the same group and because the group level mean is correlated with the unobserved group component, as we will see in the following equation.

By inverting the linear system of equations we can solve for the social interaction equilibrium level, which constitutes the first reduced form equation:

$$\mathrm{RF}_1 : \overline{Y_{gt}} = \frac{\alpha}{1-\beta} + \frac{\gamma}{1-\beta}\overline{X_{gt}} + \left(\frac{1}{1-\beta}\mu_{gt} + \frac{1}{1-\beta}\overline{\varepsilon_{gt}} \right) \tag{2.2}$$

Notice that the parameters in this equation can be consistently estimated given that the error terms have conditional mean equal to zero. However, we cannot separately identify the social interaction effect from this equation alone.

$$E\left[\mu_{gt} \mid \left\{X_{igt}\right\}_{i=1}^{N_{gt}}, G_{it} = g\right] = 0 \Rightarrow E\left[\mu_{gt} \mid \overline{X_{gt}}\right] = 0$$
$$E\left[\varepsilon_{igt} \mid \left\{X_{igt}\right\}_{i=1}^{N_{gt}}, G_{it} = g\right] = 0 \Rightarrow E\left[\overline{\varepsilon_{gt}} \mid \overline{X_{gt}}\right] = 0 \qquad (2.3)$$
$$\overline{Y_{gt}} = \pi_0 + \pi_1 \overline{X_{gt}} + \nu_{gt} \Rightarrow \begin{cases} \pi_0 = \dfrac{\alpha}{1-\beta} \\ \pi_1 = \dfrac{\gamma}{1-\beta} \end{cases}$$

By replacing the equilibrium equation in (2.1), we obtain the second reduced form equation.

$$\text{RF}_2 : Y_{igt} = \frac{\alpha}{1-\beta} + \frac{\beta\gamma}{1-\beta}\overline{X_{gt}} + \gamma X_{igt} + \left(\frac{\beta}{1-\beta}\mu_{gt} + \varepsilon_{igt} + \frac{\beta}{1-\beta}\overline{\varepsilon_{gt}}\right) \qquad (2.4)$$

This time, the exogeneity assumptions will allow us to consistently estimate the coefficients in equation (2.4) by ols. Notice that the variance covariance matrix will have a block diagonal structure, given that the error terms will be correlated across individuals in the same neighborhood, due to the presence of μ_{gt} and ε_{igt}.[32]

[32] Furthermore, assuming homoskedasticity in μ_{gt} and ε_{igt} the covariance structure will have a known form so that the equation can, in principle, be efficiently estimated by feasible generalized least squares (in a first stage β, σ_ε^2 and σ_μ^2 can be consistently estimated by ols).

$$E\left[\mu_{gt} \mid \left\{X_{igt}\right\}_{i=1}^{N_{gt}}, G_{it} = g\right] = 0 \Rightarrow E\left[\mu_{gt} \mid \overline{X_{gt}}, X_{igt}, G_{it} = g\right] = 0$$

$$E\left[\varepsilon_{igt} \mid \left\{X_{igt}\right\}_{i=1}^{N_{gt}}, G_{it} = g\right] = 0 \Rightarrow \begin{cases} E\left[\varepsilon_{igt} \mid \overline{X_{gt}}, X_{igt}, G_{it} = g\right] = 0 \\ E\left[\overline{\varepsilon_{gt}} \mid \overline{X_{gt}}, X_{igt}, G_{it} = g\right] = 0 \end{cases} \tag{2.5}$$

$$Y_{igt} = \pi_2 + \pi_3 \overline{X_{gt}} + \pi_4 X_{igt} + \omega_{igt} \Rightarrow \begin{cases} \pi_2 = \dfrac{\alpha}{1-\beta} \\ \pi_3 = \dfrac{\beta\gamma}{1-\beta} \\ \pi_4 = \gamma \end{cases}$$

Finally, a minimum distance estimator could be used to efficiently estimate the coefficients of interest from the conditions in equations (2.3) and (2.5).

An alternative to this method would be to estimate equation (2.2) by ols and use the predicted value $\widehat{\overline{Y_{gt}}} = \widehat{\pi_0} + \widehat{\pi_1}\overline{X_{gt}}$ instead of $\overline{Y_{gt}}$ in equation (2.1). By construction, $\widehat{\overline{Y_{gt}}}$ will be orthogonal to μ_{gt} and this two-stage least squares procedures will provide another consistent estimator of the social interaction effect.

B.2.- Endogenous Membership and Individual Fixed Effects

If assignment into neighborhoods is non-random, i.e. if individuals self-select into neighborhoods on the basis of characteristics of the neighborhoods (cost of living,

amenities, and characteristics of current residents) and individual preferences for the different characteristics, then the error term in the structural equation could potentially have a different distribution across neighborhoods. If this error term is correlated with the observed individual characteristics (X_{igt}), or with the aggregate measure ($\overline{X_{gt}}$), the estimator discussed in the previous section would not be consistent. We will suggest a general model of neighborhood location in which the individuals are maximizing a neighborhood specific utility, which is a function of neighborhood characteristics and individual valuations ($\delta' W_{igt}$), an individual-neighborhood effect (ϕ_{ig}) and a neighborhood specific utility shock (η_{igt}). Some components in the utility function (η and ϕ) will be correlated with the error term in the main equation (ε), giving rise to the selection problem.

$$
\begin{aligned}
&G_{it} = \arg\max_{g \in \{A,B...\}} \left\{ U_{igt} = \delta' W_{igt} + \phi_{ig} + \eta_{igt} \right\} \\
&Y_{igt} = \alpha + \beta \overline{Y_{gt}} + \gamma X_{igt} + \left(\psi_{ig} + \mu_{gt} + \varepsilon_{igt} \right) \\
&\left(\varepsilon_{igt}, \left\{ \eta_{igt} \right\}_{g \in \{A,B...\}} \right) \sim i.i.d. \\
&E\left[\mu_{gt} \mid \left\{ X_{igt} \right\}_{i=1}^{N_{gt}}, \left\{ \phi_{ig}, \psi_{ig} \right\}_{g \in \{A,B...\}}, G_{it} = g \right] \neq 0 \\
&E\left[\varepsilon_{igt} \mid \left\{ X_{igt} \right\}_{i=1}^{N_{gt}}, \left\{ \phi_{ig}, \psi_{ig} \right\}_{g \in \{A,B...\}}, G_{it} = g \right] \neq 0 \\
&E\left[\mu_{gt} \psi_{ig} \right] \neq 0 \\
&\eta_{igt} \not\perp \varepsilon_{igt}, \mu_{gt}, X_{igt}
\end{aligned}
\tag{2.6}
$$

Notice that we allowed for individual-neighborhood specific fixed effects in equation (2.6).

We are interested in understanding the conditions under which the use of individual fixed effects could eliminate the selection problem in this framework. Assuming that we have access to a two-period panel of individuals, we can express the fixed effect model as a first difference model. Equation (2.6) *for those individuals that didn't change neighborhood* between periods 1 and 2 can be rewritten as:

$$G_{i1} = G_{i2} = g \Rightarrow \Delta Y_{ig} = \beta \Delta \overline{Y_g} + \gamma \Delta X_{ig} + \left(\Delta \mu_g + \Delta \varepsilon_{ig} \right) \tag{2.7}$$

and the reduced form equations become

$$\begin{aligned} \Delta \overline{Y_g} &= \frac{\gamma}{1-\beta} \Delta \overline{X_g} + \left(\frac{1}{1-\beta} \Delta \mu_g + \frac{1}{1-\beta} \Delta \overline{\varepsilon_g} \right) \\ \Delta Y_{ig} &= \frac{\beta\gamma}{1-\beta} \Delta \overline{X_g} + \gamma \Delta X_{ig} + \left(\frac{\beta}{1-\beta} \Delta \mu_g + \Delta \varepsilon_{ig} + \frac{\beta}{1-\beta} \Delta \overline{\varepsilon_g} \right) \end{aligned} \tag{2.8}$$

with

$$\begin{aligned} \Delta Y_{ig} &\equiv Y_{ig2} - Y_{ig1} \\ \Delta \overline{X_g} &\equiv \overline{X_{g2}} - \overline{X_{g1}} \\ \Delta X_{ig} &\equiv X_{ig2} - X_{ig1} \\ \Delta \mu_g &\equiv \mu_{g2} - \mu_{g1} \\ \Delta \varepsilon_{ig} &\equiv \varepsilon_{ig2} - \varepsilon_{ig1} \\ \Delta \overline{\varepsilon_g} &\equiv \overline{\varepsilon_{g2}} - \overline{\varepsilon_{g1}} \end{aligned}$$

The first difference process eliminates the constant term and any time invariant unobserved individual and group characteristics. The first assumption we will introduce is that changes in unobserved group components are orthogonal to the changes in the average observed characteristic:

$$A_1 : E\left[\Delta\mu_g \mid \Delta\overline{X_g}, \Delta X_{ig}, G_{i1} = G_{i2} = g\right] = 0 \tag{2.9}$$

The second critical assumption is that the first difference error term has zero mean, conditional on $\Delta\overline{X_g}$, ΔX_{ig} and the location decisions on both periods:

$$A_2 : E\left[\Delta\varepsilon_{ig} \mid \Delta\overline{X_g}, \Delta X_{ig}, G_{i1} = G_{i2} = g\right] = 0 \tag{2.10}$$

To see the validity of (2.10), let's introduce the conditions implied by the location decision for the case in which there are only two neighborhoods (*A* and *B*):

$$\begin{aligned}
&E\left[\Delta\varepsilon_{ig} \mid \Delta\overline{X_g}, \Delta X_{ig}, G_{i1} = G_{i2} = A\right] \\
&= E\left[\varepsilon_{ig2} \mid \Delta\overline{X_g}, \Delta X_{ig}, G_{i2} = A\right] - E\left[\varepsilon_{ig1} \mid \Delta\overline{X_g}, \Delta X_{ig}, G_{i1} = A\right] \\
&= E\left[\varepsilon_{ig2} \mid \Delta\overline{X_g}, \Delta X_{ig}, \delta' W_{iA2} + \eta_{iA2} > \delta' W_{iB2} + \eta_{iB2}\right] \\
&- E\left[\varepsilon_{ig1} \mid \Delta\overline{X_g}, \Delta X_{ig}, \delta' W_{iA1} + \eta_{iA1} > \delta' W_{iB1} + \eta_{iB1}\right] \\
&= E\left[\varepsilon_{ig2} \mid \Delta\overline{X_g}, \Delta X_{ig}, \delta'\left(W_{iA2} - W_{iB2}\right) + \left(\eta_{iA2} - \eta_{iB2}\right) > 0\right] \\
&- E\left[\varepsilon_{ig1} \mid \Delta\overline{X_g}, \Delta X_{ig}, \delta'\left(W_{iA1} - W_{iB1}\right) + \left(\eta_{iA1} - \eta_{iB1}\right) > 0\right] \\
&= E\left[\varepsilon_{ig2} \mid \Delta\overline{X_g}, \Delta X_{ig}, \delta'\widetilde{W_{i2}} + \widetilde{\eta_{i2}} > 0\right] - E\left[\varepsilon_{ig1} \mid \Delta\overline{X_g}, \Delta X_{ig}, \delta'\widetilde{W_{i1}} + \widetilde{\eta_{i1}} > 0\right] \\
&= 0 \text{ if } \delta'\widetilde{W_{i2}} = \delta'\widetilde{W_{i1}} \text{ and the distribution of } \varepsilon \text{ remains stationary between } t = 1 \text{ and } t = 2
\end{aligned}$$

In other words, it is required that the interaction of neighborhood characteristics and individual preferences remains reasonably constant over the time periods under consideration. It is also required that the changes in the average observed characteristics are not accompanied by changes in the distribution of unobserved characteristics.

Under assumptions (2.9) and (2.10), we can estimate the reduced form equations in (2.8) and obtain consistent estimates for the parameters of interest, using the same ideas for the model with exogenous membership.

One potential problem with the previous reasoning is that some of the underlying differences across neighborhoods might actually change over time, particularly job opportunities. These differences will cause movements of individuals that might be correlated with changes in the average program or labor force participation levels. For that reason, it is important to include controls for the local economic conditions on the relevant labor markets. Furthermore, using the Instrumental Variable approach introduced in the previous section should eliminate most of the remaining endogeneity, as long as the decisions to stay or move are not systematically related to the instrument.

One might argue that the previous model is very specific in the structure of the location endogeneity. However, the basic idea is that if the location decisions of the individuals

are not drastically altered (except for idiosyncratic shocks) by the intervention that occurs between the two periods, the selectivity bias term can be "differenced out". More generally, we require that the speed of adjustment of the location process be slower than the social interaction effect.[33]

To better understand the argument, let's suppose that after the exogenous intervention occurs, the social interaction process alters the participation decisions of individuals and as a result a significant fraction of the population modify their location decision *during* t=2. The new allocation will reflect a new equilibrium of the location process (different to the one in t=1) and the new mix of individuals might have different properties in terms of the unobserved determinants of participation. Taking first differences would not eliminate the selectivity bias in this case.

Let's not forget that, even though this discussion suggests that the fixed effect approach might help eliminate the endogeneity bias, it will potentially eliminate the variation in the average participation component (or more specifically in the average value of the exogenous variable X, which acts as instrument for mean participation). For that reason, it is still important to find an intervention that differentially affects the different neighborhoods contemporaneously and over time.

[33] The role of speed of adjustment has been reported in the identification literature but never been explicitly modeled.

Appendix C: A Simple Model of Neighborhood Location and Labor Force Participation

We present a simple model that combines neighborhood choice with the decision to participate in the labor market. The idea of the model is to illustrate the fact that endogenous group formation can often lead to correlation in unobservable (and observable) characteristics among individuals in the same group. As we will see, this correlation will generate a relationship between individual and group behavior, without any social interaction in the model.

In a city consisting of 2 neighborhoods and 2 possible states (work or not work), each woman has to make 2 decisions in each period:

- Whether she will work for a fixed wage W or not work and collect welfare cash benefits for an amount of B (with $B<W$).

- Whether to live in neighborhood A or B. Depending on what neighborhood she chooses, she will have to pay rent R_A or R_B.

The individual takes these parameters as given (W, B, R_A or R_B).

These decisions will be made in order to maximize a utility function, which depends on the amount of consumption the person can afford with the income net of rent payments, the leisure time that remains from the work decision, and utility from a neighborhood specific attribute, in this case, air quality. We will assume that neighborhood *B* has good air quality and neighborhood *A* doesn't.

$$U_{itn} = U\left(C_{itn}, L_{itn}, Q_{itn}\right)$$

$$C_{itn} = \begin{cases} W - R_n & \text{if work } \left(Y_{itn} = 1\right) \\ B - R_n & \text{if not work } \left(Y_{itn} = 0\right) \end{cases}$$

$$L_{itn} = \begin{cases} 0 & \text{if work } \left(Y_{itn} = 1\right) \\ 1 & \text{if not work } \left(Y_{itn} = 0\right) \end{cases}$$

$$Q_{itn} = \begin{cases} 0 & \text{if live in neighborhood A } \left(A_{itn} = 1\right) \\ 1 & \text{if live in neighborhood B } \left(A_{itn} = 0\right) \end{cases}$$

To make the model tractable, we will assume that the utility function is additive in the different terms, according to the following specification:

$$U_{itn} = \alpha_i C_{itn} + \left(1 - \alpha_i\right) L_{itn} + \rho_i Q_{itn}$$

Where ρ_i and α_i are individual specific preference parameters independent of each other, uniformly distributed in the [0,1] interval and assumed constant in the short run.

A full general equilibrium model would require me to specify how the wage level (*W*), the welfare benefit level (*B*) and the rent differential between the neighborhoods

($R_B - R_A$) are set. For this discussion, we will assume that the system is in equilibrium and that *W*, *B* and ($R_B - R_A$) are given.

Given that neighborhood *B* has a special feature that is valued positively by all individuals, we can safely assume that in equilibrium, R_B will be higher than R_A.

Given the linearity of the problem, the two decisions can be analyzed separately. This can clearly be seen in the following table:

	Neighborhood *A*	Neighborhood *B*
Work	$\alpha(W - R_A)$	$\alpha(W - R_B) + \rho$
Not Work	$\alpha(W - R_A) + (1 - \alpha)$	$\alpha(W - R_B) + \rho + (1 - \alpha)$

Independently of the neighborhood where the individual ultimately choose to live, the work decision is given by the following condition which states that only individuals with a high preference for work relative to leisure (α) will choose to do so:

$$Y_{it} = 1 \Leftrightarrow \alpha \geq \frac{1}{W - B - 1}$$

Similarly, independently of the work decision, the location decision is given by the following condition, which implies that only individuals with a high preference for Air

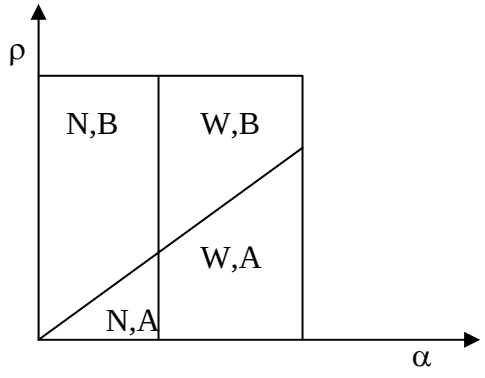
Quality (ρ) will be willing to pay the extra rent cost associated with living in neighborhood *B*.

$$A_{it} = 1 \Leftrightarrow \alpha \geq \frac{\rho}{R_B - R_A}$$

An important feature of this equation is that the (unobserved) marginal utility of consumption, α, is also part of the location decision, which means that the average level of α will be higher in neighborhood *A* than in neighborhood *B*.

The previous solution can be expressed in the following graph, showing what type of individuals would work and choose to live in the more expensive neighborhood.

The sorting equilibrium into neighborhoods and jobs (or welfare)



W = Work
N = not work
A = neighborhood A
B = neighborhood B

Given the uniformity and independence assumptions for α and ρ, the average labor force participation in each neighborhood are given by the ratios of the following areas in the previous graph:

$$\overline{Y_A} = \frac{Area(W, A)}{Area(W, A) + Area(N, A)}$$

$$\overline{Y_B} = \frac{Area(W, B)}{Area(W, B) + Area(N, B)}$$

From the previous graph we can appreciate that $\overline{Y_A} \geq \overline{Y_B}$. If an individual has a high preference for consumption relative to leisure (α), she is more likely to be living in neighborhood *A*, where average participation is higher.

This suggests that, in a model with more neighborhoods, a simple regression estimation of individual participation as a function of $\overline{Y_n}$ (the average labor force participation in the neighborhood chosen by the individual) and other determinants of work will return a positive coefficient on $\overline{Y_n}$, *without there being any social interaction present*.

Chapter 3

Information, Tax Preparation Services and Participation in the Earned Income Tax Credit

1.- Introduction

Since its expansion in the early 1990s, the Earned Income Tax Credit (EITC) has become one of the largest cash-transfer social programs in the Unites States. Given its benefit schedule (which initially increases with family earnings), the EITC is a social assistance program that provides positive work-incentives for low-income working families with children.

Despite the large benefits involved (a family with two children was entitled in 2002 to a maximum credit of $4,140) there is some evidence that not all eligible families actually claim the credit.[34] One possible explanation to this underutilization is the lack of information available to eligible families about the program rules and application procedures.[35] To claim the EITC, a family must file the regular federal income tax form

[34] Hill et al (1999) and Scholz (1994) estimated that a fraction ranging from 42% to 86% of families eligible for the Earned Income Tax Credit actually claim it. Underutilization is common to other social programs like AFDC/TANF or the Food Stamps program. Using data for the late 1980s, Blank and Ruggles (1996) estimate a 62-70% take-up rate in the AFDC program and a 54-65% rate for the Food Stamps Program.

[35] Other explanations used in the participation literature on social programs generally fall under two main categories: either families feel negatively stigmatized for receiving government aid (Moffitt (1983)) or non-participation is simply the result of a cost benefit analysis; the application costs (financial or others) exceed the benefits (Daponte et al (1999) find that information is a relevant cause for participation in the Food

and complete the information relevant to the EITC.[36] [37] As the targeted population often falls in the range of income where they are not required to file a tax return,[38] one would expect that these families might not be well familiarized with the tax system.[39] A natural source of information about how to complete a tax return and claim the EITC corresponds to a sector traditionally focused on the middle and high-earnings population: the industry of tax preparation services.[40]

In this chapter, we analyze how the decision to participate in the Earned Income Tax Credit (EITC) is affected by the presence of tax preparation services (TPS) in the

---

Stamp program but that knowledge about the program is higher when anticipated benefits are higher). Earlier studies of the EITC have found some evidence of the information hypothesis (Ross (2001) presents data from the NSAF suggesting that in 1999, only 66% of families with income below the Federal poverty level had heard about the EITC) and the cost-benefit hypothesis (Scholz (1994) finds some evidence that families eligible for higher benefits were more likely to participate). Given the private nature of the program and its focus on working families, the EITC is usually considered a non-stigmatic program.

[36] It is also possible to receive monthly payments though the employer, however only a very small fraction of the population actually use this option (less than 1%, according to Scholz (1994)).

[37] The benefits depend on the number of children (1 or two-or-more) and family earnings. The amount of the credit is deducted from the taxes owed and if the difference is negative, a check is issued for this difference.

[38] In 1999 the minimum amounts required to file a tax return were $9,100 and $12,700 for taxpayers filing as “head of household” or “married filling jointly”, respectively.

[39] Hill et al (1999) find that most of the non-participant eligible families were families that didn’t file a tax return.

[40] Another possible source of information is the “word of mouth” among friends, colleagues or neighbors. A planned extension for this article is to quantify the importance of information diffusion at the neighborhood level for EITC participation.

neighborhood.[41] Conceptually, the presence of tax preparation services would provide both an informative role as well as reducing the transaction cost of participation.

In effect, there is some evidence that these services may have played a significant role in the relatively high penetration of the program among low-income families.[42] This study will formally quantify this hypothesis, by analyzing individual participation among welfare-prone families in California and how it relates to the presence of TPS in the zip code of residence. To do so, we combine administrative data of welfare recipients in California and information from the Employment Development Department (EDD) about the number of tax preparation units per zip code between 1993 and 1999.

The main concern with this analysis is the potentially endogenous location of tax preparation services in areas with higher potential for EITC participation (areas with more potentially eligible families). If this is the case, a correlation between EITC participation and presence of tax preparation services could simply reflect that these firms are doing a good job at anticipating where EITC eligibility is growing. In fact, some evidence suggests that tax preparation services were opening offices in low-income

[41] To our knowledge, there is only one article that has analyzed the relationship between tax preparation services and the EITC. Berube et al (2002) focus on the cost associated with the advanced loans provided by tax preparation services and suggest that an increasing share of their business came from this lucrative service.

[42] Hill et al (1999) finds that up to 68% of the low-income families in their sample that filed a tax return used a tax preparation service. According to Berube et al. (2002), approximately half of the 1999 EITC refunds were transferred through advance loan tax preparation services.

neighborhoods with the specific goal of offering (high-interest) advance loans for EITC eligible families.[43]

To formally study the extent of this problem, we use 1990 census data to construct zip code level measures of potential EITC eligibility over time. Variation in these measures over time captures changes in the benefit structure (particularly the expansion of benefits) and the initial 1990 socio-economic structure.[44] We then estimate how these measures can predict the variation over time in the availability of tax preparation services. Our findings on the location endogeneity suggest that, at least for California and contrary to popular belief, TPS tended to locate more in areas with lower EITC eligibility potential. This could be consistent with the idea that these services were more interested in exploiting the increase in traditional tax preparation demand (non EITC related) that resulted from the economic expansion. Traditional customers have higher incomes and thus low EITC eligibility.

Our main findings concerning the impact of tax preparation services on EITC participation suggest that increasing the number of TPS by one unit (the average number is 2.16) in an averagely sized zip code would increase the likelihood of participating in the EITC by approximately 10 percentage points. Given our earlier findings about the

[43] Berube et al. (2002) analyze an IRS database on Electronic Return Originators (EROs) suggesting that the concentration of EROs per filer is much higher in high EITC filing zip codes. These results are not conclusive of the existence of the endogeneity problem, however, given that they only look at EROs (therefore loosing the large number of paper-only services), that the denominator corresponds to the total number of filers (and as we mentioned earlier, non-participation seems to be associated with non-filing) and that they only look at large metropolitan areas for one year of data (1998). A more extended discussion is presented in Section 2.

negatively endogenous location, we can interpret our estimates as lower bounds on the true impact (had the TPS located more evenly across the EITC eligibility distribution, the impacts may have been larger). Our findings about the effect of TPS on the incidence of filing a tax return (independently of the EITC status) are very similar, consistent with previous research, which suggested that EITC non-participation is directly related with not filing a tax return (rather than filing a tax return and failing to claim the EITC).[45]

The rest of the article is divided as follows: In section 2, we present background information about the Earned Income Tax Credit, and we analyze the location decision of TPS. In section 3, we present the empirical models to be estimated and we describe the different data sources used for estimation. In section 4, we present the results and in section 5, we conclude.

2.- Background

2.1.- The Earned income Tax Credit[46]

The Earned Income Tax Credit is a Federal Program created in 1975 that provides a subsidy for low-income families with children. Initially small, the program was greatly expanded in the early 1990s to become the largest cash transfer program in the US.

---

[44] Arguably, the 1990 Census is practically the only reliable data on zip code level demographics, so we would expect that it would also be used by TPS, when making their location decisions.

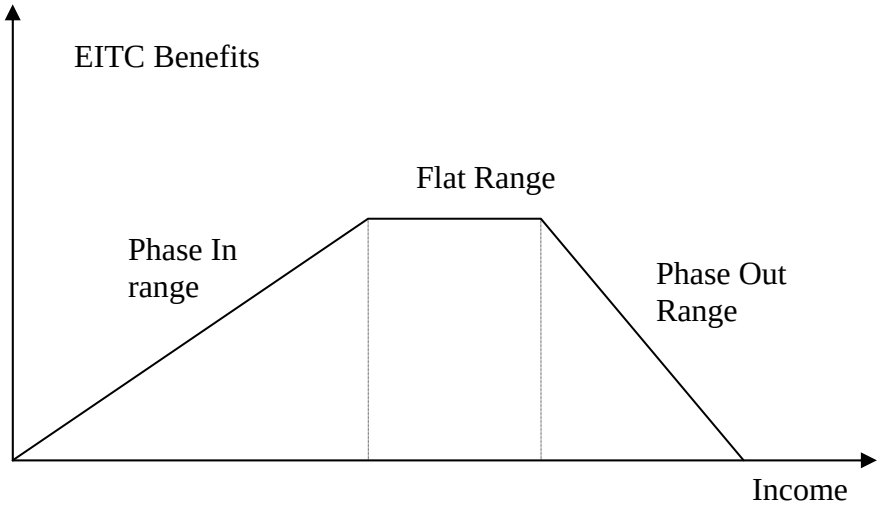
[45] In our sample, 92% of families filing a tax return also claimed the credit. The remaining might include families who are in fact not eligible for the program, because of high income or failing to pass relationship tests.

[46] See Hotz and Scholz (2000) for an extensive review of the EITC, its history and behavioral effects.

The program is operated through the income tax system; a family (either a married couple or a head of household) with children that had some earned income during a year can file, even if they do not owe any taxes, an income tax form claiming the credit. The benefit works as a refundable credit: if the amount of the benefit is higher than the tax liability, a check is issued for the difference.

In Figure 1, we show the benefit schedule of the program. Starting at zero earnings, the level of benefits increases with earnings, up to a certain point. This is called the "phase-in range" of the program. Benefits remain constant during the "flat range" and are then reduced in the "phase-out" range.

Figure 1 – EITC Benefit schedule

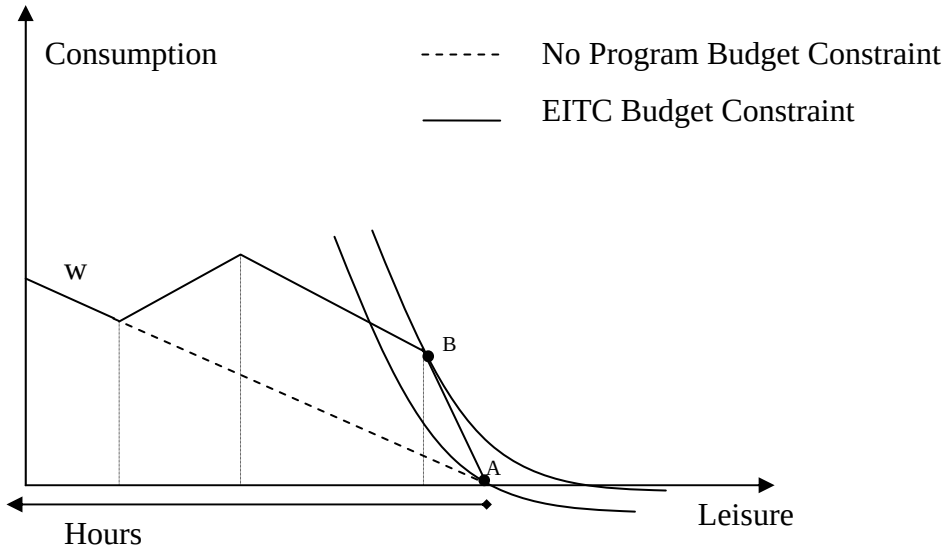


In contrast with traditional social programs, the EITC provides an unambiguous incentive to participate in the labor market, at least for single headed households. This can be seen in Figure 2, where we show the traditional consumption-leisure trade-off model,

comparing the regular budget constraint (without EITC) and the modified budget constraint.

In absence of the program, an individual with strong preference for leisure relative to consumption would choose not to work (point A). With the introduction of the program, the effective wage faced by an individual in A is increased by the phase-in credit rate, making the individual to choose a positive work supply (point B).

Figure 2 – Work incentive effect of the EITC



As we can see in table 1, the benefits schedule depends significantly on whether an eligible family has only one child or two or more children. Both the Phase-In credit rate (the slope of the line in the phase-in range) and the maximum benefits (maximum credit) were expanded between 1993 and 1999, but especially for families with two or more children.

As the current literature evaluating this pro-work incentive effects has consistently found positive effects of the program on labor force participation, some States have implemented or are considering implementing their own versions.

Two additional characteristics should be mentioned about the program, which separate it from traditional welfare programs: the program does not distinguish between single parents or married couples and the benefits are capped for family size, i.e. the amount is the same for families with 2 or more children.

2.2.- Participation in the Earned Income Tax Credit

The more recent study on EITC participation corresponds to Hill, Hotz, Mullin and Scholz (1999), who analyze data from the 1993-1994 welfare caseloads in 4 Counties in California. The most relevant findings for the current study are the following:[47]

- Between 42% and 84% of eligible households actually claimed the credit.[48]
- Non participation is usually coincident with not filing a tax return.
- Between 55% and 68% of the families filing a tax return used a tax preparation service.

Notice that one of the goals of that study was to analyze the take-up rate of the program, i.e. the fraction of eligible families who participate. This implied restricting the sample to families with children and positive earnings. In this article, we selected cases that were on

---

[47] The basic data we use in this article comes from the same source as Hill et al. (1999).

[48] The range of results comes from different assumptions about earnings and the family structure. As in our case, no detail is provided about the exact relationship between persons in the same welfare case.

welfare at some point, independently of their earnings. The idea behind this choice is that if the program has the pro-work incentive attributed to it, work and participation will be to some extent a joint decision so that even a family with no earnings is considered as "potentially eligible".

In an earlier study, Scholz (1994) used matched data from the Survey of Income and Program Participation (SIPP) and federal tax returns to estimate participation (take-up) in the EITC and explore the determinants of this decision. His "preferred" estimates of participation are between 80% and 86% in 1990. According to the author, the findings from a probit model of EITC participation among eligible families are "consistent with voluntary or rational explanations for nonparticipation". This article does not have information about the presence or use of tax preparation services.

In a more recent analysis, Berube et al (2002) document the usage of tax preparation services in the 100 largest metropolitan areas, with special focus on the costs associated to preparing, filing and providing advance loans to eligible families. They also present some evidence on the concentration of Electronic Return Originators (EROs) in neighborhoods with different levels of EITC claiming rates (calculated as the fraction of tax returns that claimed the credit in a given zip code). The results most relevant for the current article are summarized as follows:

- Taxpayers claiming the EITC and using an ERO paid in average around $204 in preparation, filing and the advanced loan feature (more than 10% of the EITC benefit).
- The largest commercial tax preparation service earnings from "fast cash" products went from $138 billion in 1998 to $357 billion in 2001.
- EROs concentration is positively correlated with EITC claiming rate.
- In 1999, approximately half of the total EITC claimed ($30 billion) was refunded through advanced loans.

The most relevant finding for the current study is the idea that TPS might endogenously locate in areas with high EITC participation, generating an identification problem for the interpretation of our results: Do TPS increase participation or does higher eligibility attracts new offices to locate in an area?

These results, however, are limited in at least three dimensions:

- The focus on EROs is a biased measure of the availability of tax preparation services. The authors argue that EITC eligible low-income families are more likely to use EROs, based on the idea that these families are in higher need of "fast cash" and that EROs are the main providers of this type of service. No hard evidence is however supplied to sustain these claims and historically, a large majority of all returns are filed in paper form and sent to the IRS.[49]

[49] In 1997, 78% of all returns were "paper returns" (own calculations based on IRS, Statistics of Income Bulletin, Publication 1136, summer 1998 and Taxpayer Usage Study published weekly from February 1998 to April 1998). Electronic filing has been increasing, however, and in fiscal year 2002, it represented close to 36% of all individual income tax returns.

- The denominator used in the concentration analysis corresponds to the total number of tax returns filed in the area. As previous research suggests that non-participation is associated with not-filing a tax return, the fraction of EITC filers does not really represent participation among the eligible population. Therefore, if high EITC claiming (relative to total number of returns) areas are also areas with low filing rates, the previous relationship could be reversed.
- The analysis is restricted to the 100 largest metropolitan areas in 1998.

As we can see, existing research seems to reinforce the idea that tax preparation services are likely to play an important role in the penetration of the program among the welfare-prone population. In the next subsection, we will analyze more carefully the issue of endogenous location of TPS in high eligibility areas.

2.3.- Tax Preparation Services

In this section we will provide evidence that, contrary to popular belief, tax preparation services (TPS) were not endogenously locating in low-income areas to target the EITC population. In fact, the evidence presented suggests that the expansion in the availability of tax preparation services in California was more likely to follow the increased potential for services in traditional middle or high-income areas.

In general, any resident of the US with a minimum level of earnings must complete a federal tax form and, in some cases, a State income tax form. As both average and

marginal tax rates generally increase with earnings, the incentives to hire a professional tax preparer will also increase with earnings; professionals can provide advice on measures to reduce the tax burden, by itemizing deductions, shifting assets, etc. For this reason, the traditional customers of TPS were middle and high-income taxpayers, for which the willingness to pay for this service is high enough to cover its cost.

The business potential of low-income working families comes from their relative lack of experience with filing income tax forms and the high one-time benefit that can be obtained by claiming the EITC (and therefore the high willingness to pay for being able to do so). In order to claim the EITC, an eligible family has to:

- Know about its eligibility and potential benefits
- Complete a regular income tax form plus the additional information specific to the EITC (Schedule EIC) and submit it (via mail or electronically).

If a family has a level of income that requires them to file a tax return, the marginal cost of claiming the EITC will be small, relative to the case where the only reason to complete the entire tax form is to claim the credit.[50] Tax preparation services (TPS) can significantly help reduce this burden by their knowledge of the rules, the ability to easily fill a form and send it to the relevant authority. They can even provide the additional

[50] The Gross Income level above which a person filing as head of household under age 65 is required to file a tax return went from $7,550 in 1992 to $9,100 in 1999. These thresholds are very close to the beginning of the flat region of the EITC, where the maximum benefits are paid (see Table 1).

service of offering an advanced loan for the expected refund.[51] As the amount of the credit is paid in one lump sum payment as big as $3,816 for a family with 2 children in 1999, the cost of paying a tax preparation service can be easily covered by the benefits.[52]

This is one of the reasons why anecdotal evidence suggested that TPS had the incentives to target the low-income population by opening offices in traditionally neglected areas. As we discussed earlier, the evidence presented by Berube et al (2002) is weakened by the type of data used in their calculation and, in any case, it only provides evidence for the last part of the 1990s.

To explore more formally whether TPS location decisions during the 1990s were more affected by the EITC potential or the increasing demand from the traditional sector, we used zip code level data on the number of tax preparation services operating during the first quarter of every year. The source of this data is the department in charge of managing the unemployment insurance program in California (Employment Development Department, EDD). All employers send quarterly reports about the wages paid to their employees to EDD, and each firm is classified into an industrial sector. As there is unique code for "Tax Return Preparation Services" (Industry code 7291), it was possible to obtain information on how many firms were operating in every zip code, the

[51] Evidence suggests that the implicit interest rate behind these advanced loans was quite high. This called the attention of the press and motivated articles like Berube et al (2002).

[52] Berube et al (2002) estimated that the average payment to TPS for filing and providing an advance loan was approximately $200.

number of employees and the average quarterly earnings of these employees. For confidentiality reasons, however, employment and earnings data cannot be reported when too few firms are operating in an area or when one firm concentrates more than 80% of the employment in the sector. For that reason, we restricted our analysis to the information on number of firms operating. We chose the first quarter of every year, as this represents the majority of the tax season and this is an industry marked by the seasonality of its employment.

Figure 3 presents the evolution of the total number of tax preparation services operating in California between 1991 and 1999, together with the State yearly unemployment rate and the maximum EITC benefit for a family with 2 children.[53] Apparently, some of the expansion in the number of TPS could be explained by the economic condition of the State; the number of TPS remained stable between 1991 and 1993, which coincides with the period of high unemployment rate and began to increase as the economy recovered, reaching a maximum in 1999, when the unemployment rate was at its minimum. At the same time, EITC benefits increased in real terms between 1993 and 1996 and remained stable after that. Obviously, we cannot distinguish from this graph which of the two effects (the EITC or the economic expansions) is more likely causing the expansion in the number of TPS providers.

[53] Source for unemployment rate: Labor Market Information Division of California Employment Development Department.

To specifically test whether the expansion of these services responded to the increased eligibility potential of the EITC, we match the information on the number of tax preparation services by zip code with data from the 1990 census and construct 4 measures of the EITC potential over time:

- **EITC Potential 1:** We obtain the number of families for each earnings range (0-5000, 5000-10000, etc.) from the 1990 Census and assign the maximum 2-children real EITC benefit (in 1999 US$) for a family in the midpoint of each income bin ($2500, $7500, etc.). These benefits are added over all income ranges eligible for the EITC. As the EITC benefits schedule was changing over time, this measure will also be changing and in a different way for each neighbourhood (as the initial distribution will be different across areas).

$$EITC^{1}_{potential} = (\#\ families \in [0;5000]) * (Max\ EITC\ 2children(2500)) + (\#\ families \in [5000;10000]) * (Max\ EITC\ 2children(7500)) + \ldots$$

- **EITC Potential 2**: Similar to **EITC Potential 1** but using the 1-child maximum EITC benefits.
- **EITC Potential 3**: We obtain the number of families with children and earnings below the poverty level from the 1990 Census and assign the 2-children maximum real EITC benefit for a family in the flat range.

$$EITC^{3}_{pot} = (\#\ families\ w/children,\ in\ poverty) * (Max\ EITC\ 2children)$$

- **EITC Potential 4**: Similar to **EITC Potential 3** but using the 1-child maximum EITC benefits.

The first two measures allow us to use a relatively fine detail on the earnings distribution of families across zip codes but include both families with and without children. Measures 3 and 4 restrict the sum over poor families with children but do not allow finer detail on the income level of families. Unfortunately, there are no publicly available measures of the joint distribution of earnings and number of children at the zip code level.

We then estimate regression models where we used the number of tax preparation services (and whether this number was greater than zero) as a function of the different EITC eligibility potential measures, year dummy variables and zip code level fixed effects.

$$TPS_{zt} = \alpha + \beta\left(EITC_{pot(zt)}\right) + \eta_z + \varphi_t + \varepsilon_{gt}$$
$$\left(TPS_{zt} > 0\right) = \alpha + \beta\left(EITC_{pot(zt)}\right) + \eta_z + \varphi_t + \varepsilon_{gt}$$

where TPS_{zt} represents the number of Tax Preparation Units in zip code *z*, year *t*.

The time dummies eliminate the secular trend observed for California as a whole and the zip code fixed effect control for time invariant characteristics that could attract TPS services to some neighborhoods and not to others. Identification of the EITC measures comes from the variation over time in the benefits available for individuals with different income levels: zip codes with higher proportions of families in the flat range of the EITC (where benefits are maximal) will experience higher increases in the demand for EITC related tax preparation services. As the demographic distribution of zip codes is taken from a period before the EITC expansion (reported earnings in the 1990 Census

correspond to 1989 earnings) and it is kept fixed for the entire period, we are confident that our results do not reflect a reverse causation problem: that the presence of TPS was causing changes in the distribution of eligible individuals across neighborhoods. If the coefficient on the eligibility potential is positive, it means that tax preparation services were in fact reacting to this increased EITC potential. On the contrary, as the eligibility measures reflect a combination of the fraction of low income individuals in a neighborhood and the EITC expansion (which increased at the same time that the economy was improving), a negative coefficient would reflect that the TPS industry expanded in relatively well-off neighborhoods, away from the EITC eligible population.

The results of this set of models are presented in table 10, were each column represents a regression. As unreported results suggested that there might be some non-linearities in the relationship, we also included models where the EITC measures were interacted with the position relative to the median in the population distribution of the zip code: a "small zip" has a population below the median (22,090 individuals) and a "large zip" is located above the median.

The results can be summarized as follows:

- In all basic specifications, the EITC eligibility potential has a negative sign, statistically significant at 5% in all but 1 specification.

- The effect of EITC eligibility on whether there is at least one TPS in the neighborhood are slightly smaller than the effect on the total number of neighborhoods.
- The effect varied significantly by the size of the zip code: in largely populated zip codes, the effect is very similar to the overall effect (negative and statistically significant) whereas in smaller neighbourhoods, the effects are positive for the existence of TPS but small and statistically insignificant.

Overall, the results suggest that the expansion in TPS services occurred mostly in largely populated zip codes, where some TPS were already operating. Some evidence exists of a small expansion into low-income neighbourhoods where no TPS were previously operating.

With these results in mind, we feel more confident that if we find evidence that the availability of tax preparation services increased the likelihood of participation in the EITC, the results will not be driven by the endogenous location choices of TPS.

3.- Empirical specification and data

3.1.- Data Sources and sample construction

Despite the importance of the EITC in terms of expenditure, number of participants and its theoretical pro-work effects, the research in the area has been affected by the lack of data sets that actually include a reliable measure of program participation. As the program is administered through the tax system, its confidentiality is secured with the

highest standards. Even if traditional surveys (like the Consumer Population Survey (CPS) or the Survey of Income and Program Participation (SIPP)) have regularly included questions about the EITC, researchers have been reluctant to exploit this information, probably because of low response rates or lack of confidence about the accuracy of the answers.

In this work, we will use a unique source of information about the EITC.[54] The base data set consists of the universe of individuals that participated on the AFDC/TANF program at some point between 1987 and 1999 in California. This sample is originated in the California Department of Social Services MediCal Eligibility Data System (MEDS) which records every individual eligible for the state Medicaid program. The MEDS file records the particular aid code under which an individual was eligible for MediCal (including the Family Unit (FG) or Unemployed Parent (UP) status) for every month in which the person was eligible. It also includes the sex, age, ethnic code and language and in this particular confidential version, the ZIP code where the person lived.

It is important to notice that the individual information is collected only while the person is eligible for MediCal. This implies that it is not possible to perfectly follow an individual (in particular, its location) when she is not eligible for MediCal.

The MEDS sample was matched with quarterly earnings data from the department in charge of the Unemployment Insurance in California (EDD) for the whole period

(whether the person was on welfare or not). This source includes earnings from jobs subject to the Unemployment Insurance program and thus, excludes both government jobs as well as self-employed earnings.

Finally, and following a very strict security arrangement, the selected sample is sent to the State tax authority (the California Franchise Tax Board (FTB)) where a new matching process is performed and the EITC participation (more precisely, whether the person claimed the EITC) is appended to the data set. The FTB then executes the necessary statistical analysis and after following a disclosure analysis, the estimation results are forwarded to the researcher.

Using the information on the zip code and county of residence of each family, we matched information about the number of tax preparation services operating in the zip code (see discussion in previous section) and aggregate measures at the county level on the employment to population ratios for all sectors. The latter measures were constructed by combining total employment data (average over the 4 quarters) from the California Employment Development Department (EDD) and yearly population estimates from the California Department of Finance. As mentioned earlier, EDD collects data from all employers on the earnings paid to each employee and produces quarterly employment and earnings statistics by industrial sector at the county level.

[54] Hill et al. (1999) were the first to look at EITC participation using practically the same sources of data,

For the empirical analysis of this article, we created a series of cross sections of families that were receiving AFDC/TANF between 1993 and 1999, keeping track of family characteristics and observing their earnings, whether they filed a tax return, whether they claimed the EITC, and the zip code and county of residence to assign labor demand and TPS availability. As we wanted to explore the effect of tax preparation services over a long period of time and given that geographic location of families is only observed while receiving welfare, we decided to lose the panel dimension of the data and concentrate on constructing representative samples of welfare-prone families for each zip code at every point in time. This explains our choice of repeated cross-sections of individuals. At the same time, we didn't want to force the individuals to be necessarily receiving aid during the analysis year, which explained why we extracted families that were on welfare during the last quarter of the year previous to the analysis year. We implicitly assumed that these families where still residing in that zip code during the tax season of the following year. This could cause an upward bias in our estimate of the impact of TPS availability, only if there is a large fraction of families who are changing zip codes from year to year and that this mobility is positively related to the number of TPS: for this to happen, it would have to be the case that families living in neighborhoods with a high number of TPS and who decide to move, experience a high "jump" in the number of TPS, relative to those families who live in low-TPS neighborhoods.[55]

The timing for the different variables is as follows:

---

but for a sample of 4 counties in California.

- We observe all families with children that were receiving aid in California during the last quarter of a particular year (for example October-December 1992). At this point, we construct basic family characteristics: number of children, how many children under age 1, under age 5, whether it is a single parent case (Family Group, FG) or a two-parents case (Unemployed Parent, UP) and the age and race of the oldest person. We also obtain the residential location, given by the 5-digit zip code and county of residence.
- This will represent the population of welfare prone families during the following year (the *analysis year*, 1993 in the previous example).
- For each family, we observe the total earnings from the adults in the case during the *analysis year*, as well as whether any adult in the case filed a tax return for the earnings received during the analysis year and claimed the EITC (these tax returns were collected during the year following the analysis year, 1994 in the previous example). From individual filing and participation measures, we construct case level measures: a family filed a tax return if any person in the case filed a return either as the taxpayer or as the husband or wife of a taxpayer (similarly for EITC participation).
- We merged information on the number of tax preparation services operating in the zip code, during the tax season corresponding to the *analysis year* (first quarter of 1994).

[55] In future work, we will study in more detail the determinants of migration of welfare recipients.

- We also merged information on labor market conditions, proxied by the average employment to population ratio at the county level, during the *analysis year*.

3.2.- Empirical models

The goal of this paper is to analyze the pure effect of the availability of TPS at the zip code level on the EITC participation of welfare-prone families residing in the area. We do so by estimating linear models where the dependent variable is 1 if the family filed a tax return or claimed the EITC and the independent variables are:

- The availability of TPS in the zip code, measured as the number of tax preparation services, divided by the 2000 Census estimated population for that zip code.
- The maximum EITC benefit available to each family, which is only a function of the number of children and the year of analysis.
- A labor demand measure, the county average employment to population ratio (EP) for the year.
- Family characteristics (X): The number of children under age 1, under age 5, the age of the oldest person, the squared age, a dummy variable equal to 1 if the case was a Family Group case (1 adult) or 0 if the family was an Unemployed Parent case (2 adults), and race dummy variables White, Black or Hispanic (the default group is either Asian or other races).

The first specification is a linear model on the independent variables, which will be estimated by Ordinary Least Squares (OLS):

1) OLS: $$Y_{iczt} = \alpha + \beta(TPS_{zt}) + \chi(EITC_{iczt}) + \delta(EP_{ct}) + \gamma' X_{iczt} + \varepsilon_{icgt}$$

where Y_{iczt} corresponds to the outcome of interest (Claimed the EITC or filed a tax return) for family *i*, leaving in county *c*, zip code *z*, during analysis year *t*.

In order to account for (time invariant) unobserved differences at the zip code level that could be correlated with the TPS variable and affect EITC participation, we estimate a specification that also includes a zip code fixed effect:

2) Zip FE: $$Y_{iczt} = \alpha + \beta(TPS_{zt}) + \chi(EITC_{iczt}) + \delta(EP_{ct}) + \gamma' X_{iczt} + \eta_z + \varepsilon_{icgt}$$

3.3.- Summary Statistics

In table 9, we present the summary statistics for the analysis sample.

As expected, most of the caseload (84%) is composed by single parent families (FG) and the average number of kids under age 1 is 0.91. The average age of the oldest adult is 32.6, and the population is 32% Hispanic, 31% white and 23% black (14% of other races). The average real maximum benefit these families are potentially eligible for is $2,906.

In this sample, only 53% of the cases had positive earnings among the adults, 43% filed a tax return and 40% claimed the EITC. We also report some conditional probabilities: conditional on filing a tax return, a large proportion of the cases claimed the EITC (92%) but only 70% of the cases with positive EDD earnings actually filed a tax return. This is consistent with previous findings that the question of participation in the EITC is essentially a question of filing a tax return.

We also report a take-up measure, the fraction of families who claimed the EITC among those that worked (that had positive EDD earnings).[56] We found that during this period, 64% of eligible families actually claimed the credit.[57] This estimate falls in the range estimated by Hill et al. (1999) and slightly below that of Scholz (1994).

In our sample, the average number of individuals per zip code is 991 but with high variation.[58]

Finally, the average number of TPS in a zip code is 2.16, with many zip codes having none.

[56] All families in the sample have children.

[57] This is a raw measure of take-up, as we did not verify whether earnings were below the upper EITC threshold or whether the children were present for at least 6 months.

[58] In order to reduce the errors associated with invalid zip codes, we restricted ourselves to those zip codes with at least 50 sampled individuals.

4.- Results

4.1.- Claiming the EITC

In Table 11, we present the results from the 4 specifications presented in the previous chapter: two specifications for EITC participation and two specifications for the tax filing decision. Each column represents a different specification, following the empirical models described in section 3.2.

The OLS coefficient (4.6) in the first specification can be interpreted by saying that increasing the number of TPS in one unit in an averagely sized zip code (with a population of 46,000 individuals in 2000) would increase the likelihood of participating in the EITC in approximately 10 percentage points (4.6 / 46).

Except for the age variable, all other covariates have significant coefficients and of the expected sign: increasing the maximum EITC benefit by a thousand 1999 US$ increases participation by approximately 9.5 percentage points. Better economic conditions at the county level also increase participation. Being a single parent or having kids under age 1 or 5 tend to reduce participation. Finally, black or Hispanic families tend to have higher participation than whites or other races.

Remarkably, when introducing zip code fixed effects, the impact of higher TPS availability increases significantly (to 16.1). One interpretation for this change is that the OLS estimate exploits both the variation within zip codes (over time) and between zip

codes, whereas the fixed effect model relies on the within variation only. The large difference between the two estimates seems to reflect that most of the effect of tax preparation services occurred through the expansion over time in their availability, rather than the fact that some neighborhoods had, in general, more TPS than others.

The third and fourth columns correspond to the equivalent results for the decision to file a tax return. The general pattern is very similar to the previous results: we find a strong positive effect for the availability of TPS in the neighborhood. The other covariates included in the regression have the predicted signs and significance.

5.- Conclusions

In this article, we have explored the importance of information in the participation decision of one of the main welfare innovations of the last decades in the United States. We looked at a formal channel of information transmission (the availability of tax preparation services in the neighborhood). Our results suggest strong effects for the presence of TPS in the zip code, implying that private (for profit) institutions could be of valuable help in disseminating information about a new social program.

This is in part possible because the program benefits are of a significant magnitude and are paid in one lump sum payment (as opposed to monthly payments). That could in part explain why there aren't similar institutions operating in other welfare programs (like TANF or Food Stamps). Some of the current research on program participation has advocated that government should subsidize outreach efforts in order to improve take up rates in government programs.[59] Further research should be pursued to investigate whether reducing the frequency of payments (and thus increasing the amount paid each time) would generate the incentives for the private sector to naturally provide application assistance, without the need of an additional subsidy. Private assistance can be the defining factor that allows a program to reach the relatively high take-up rates observed in the Earned Income Tax Credit.

[59] Aizer (2001) and Aizer (2003) investigated the effect of contracting out outreach to private organizations or financial incentives to community based organizations in order to improve health insurance coverage among low income children. Her findings reinforce the idea that information can be a significant barrier to program participation and that private outreach institutions can help overcome these barriers.

Figure 1 – EITC Benefit schedule

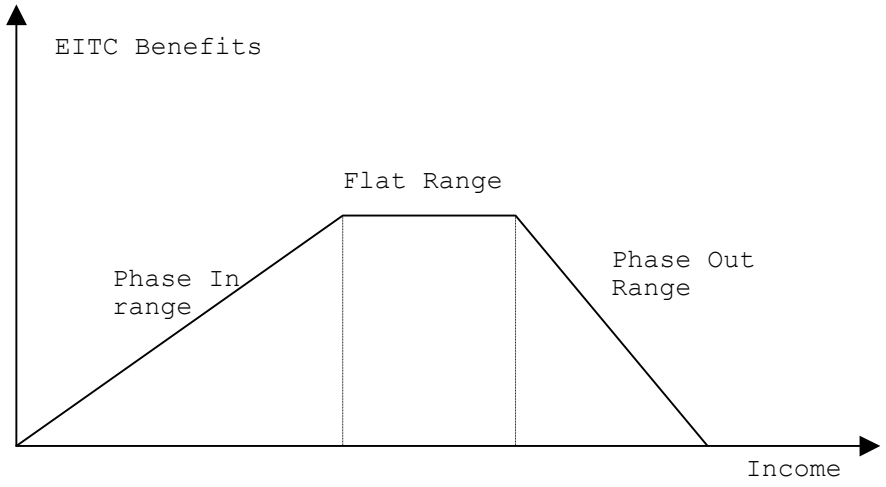


Figure 2 – The work incentive of the EITC

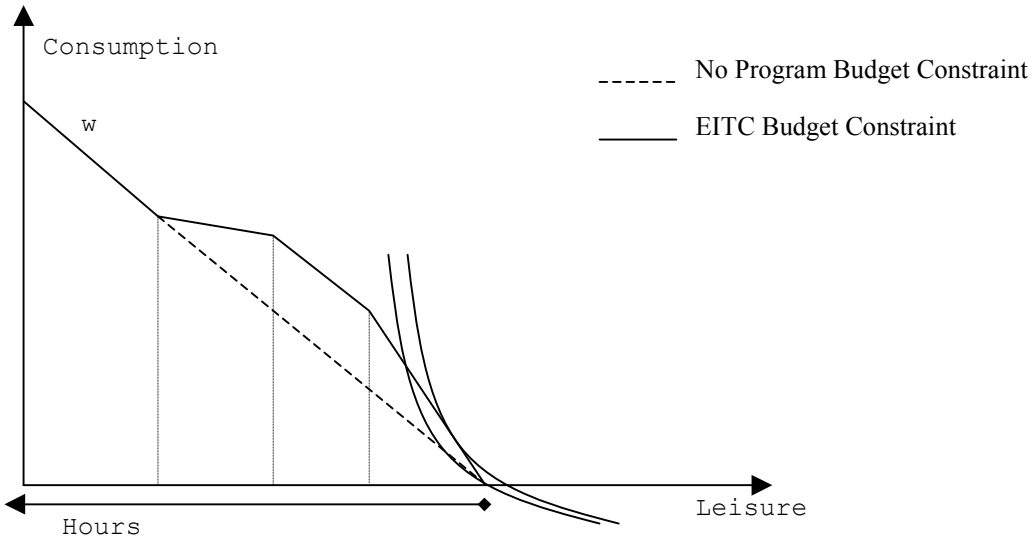


Figure 3. Distribution of average change in LFP, zip code, 93-94

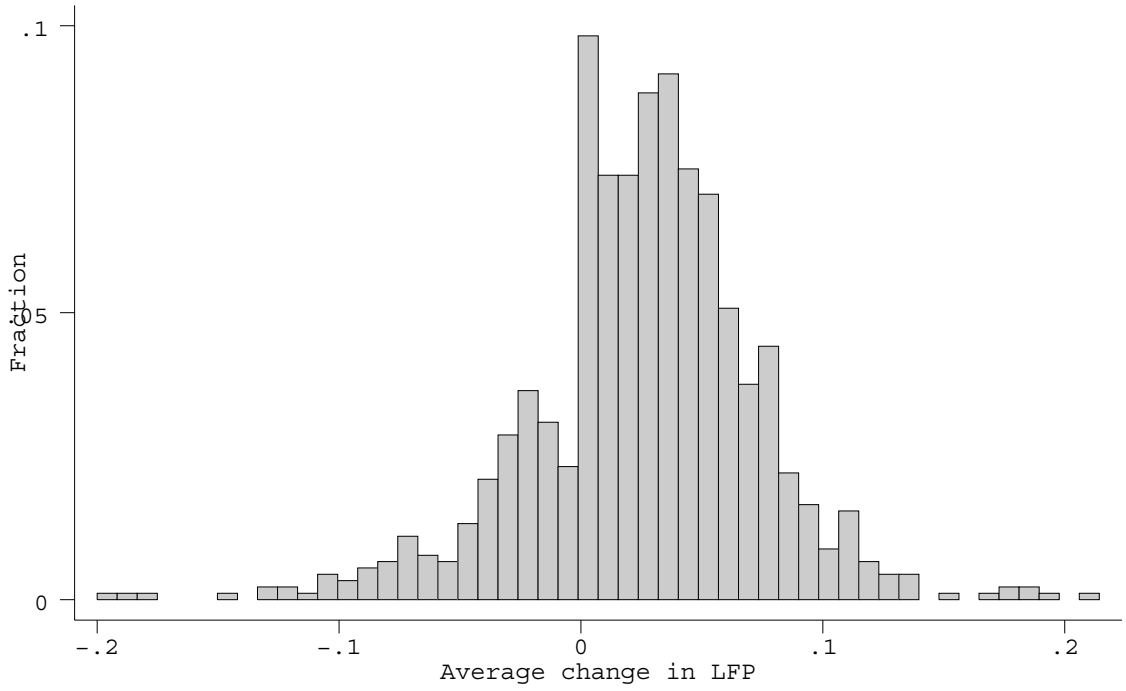


Figure 4. Distribution of average fraction 1 vs 2+ chidren, 93-94

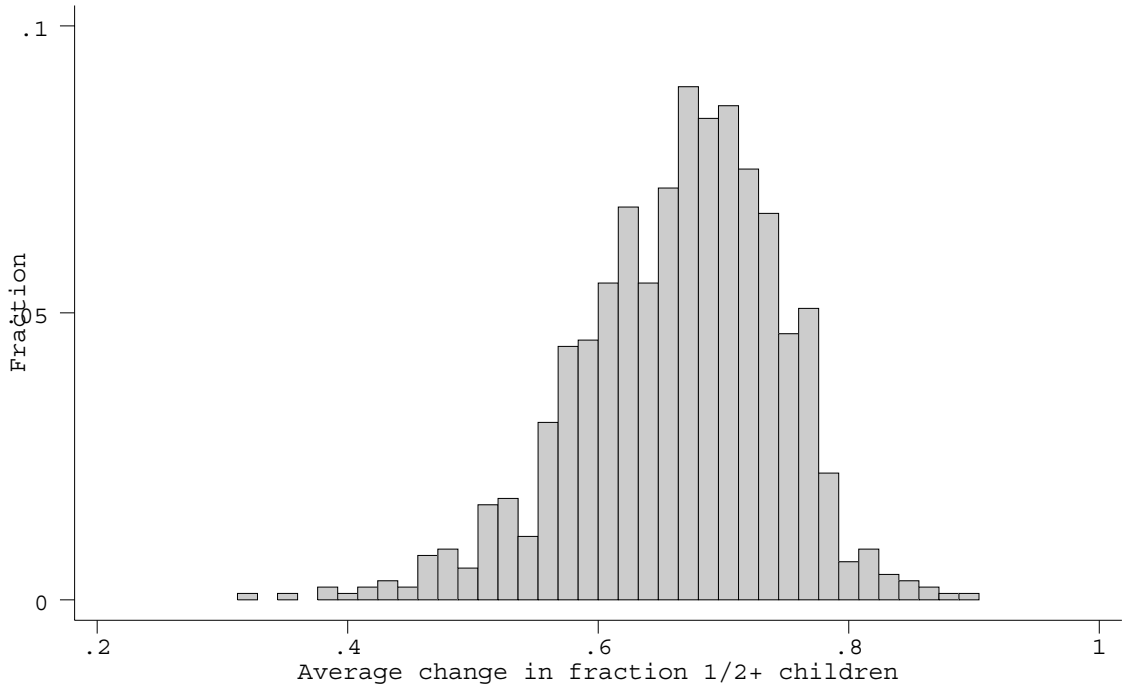


Figure 5 - Tax Preparation Services, Unemployment and Maximum EITC Benefits (family with 2 children)

Table 1 – EITC parameters and Gross Income above which families are required to file a tax return (1993 – 1999)

	Phase-In Region	Flat Region			Phase-Out Region		Minimum Gross Income
Year	Credit Rate	Begining	Ending	Maximum Credit	Phase-Out Credit rate	Maximum Income	Head of Household
1993							
1 child	18.5%	$7,750	$12,200	$1,434	13.21%	$23,050	$7,550
2+ children	19.5%	$7,750	$12,200	$1,511	13.93%	$23,050	
1994							
1 child	26.30%	$7,750	$11,000	$2,038	15.98%	$23,755	$7,800
2+ children	30.0%	$8,425	$11,000	$2,528	17.68%	$25,296	
1995							
1 child	34.0%	$6,160	$11,290	$2,094	15.98%	$24,396	$8,050
2+ children	36.0%	$8,640	$11,290	$3,110	20.22%	$26,673	
1996							
1 child	34.0%	$6,330	$11,610	$2,152	15.98%	$25,078	$8,250
2+ children	40.0%	$8,890	$11,610	$3,556	21.06%	$28,495	
1997							
1 child	34.0%	$6,500	$11,930	$2,210	15.98%	$25,750	$8,450
2+ children	40.0%	$9,140	$11,930	$3,656	21.06%	$29,290	
1998							
1 child	34.0%	$6,680	$12,260	$2,271	15.98%	$26,473	$8,700
2+ children	40.0%	$9,390	$12,260	$3,756	21.06%	$30,095	
1999							
1 child	34.0%	$6,800	$12,460	$2,312	15.98%	$26,928	$8,950
2+ children	40.0%	$9,540	$12,460	$3,816	21.06%	$30,580	

Source: Hotz and Scholz (2002) and IRS Instructions Form 1040, various issues.

Table 2 - Summary Statistics - Chapter 2

	1993		1994	
	mean	s.e.	mean	s.e.
Families				
Number of observations	142391		142391	
Labor Force Participation	0.337	0.473	0.367	0.482
Number of children	2.220	1.120	2.186	1.091
More than 1 child	0.691	0.462	0.691	0.462
Mother of white race	0.288	0.453		
Mother of black race	0.195	0.396		
Mother of Hispanic origin	0.428	0.495		
Mother's age in 1994	33.89	5.95		
County Employment / Population, all sectors	0.603	0.091	0.610	0.091
County Employment / Population, retail sector	0.104	0.010	0.104	0.010
County Employment / Population, services sector	0.161	0.046	0.164	0.046
Unemployment rate	10.27	3.32	9.44	3.04
Zip codes				
Number of zip codes	503		503	
Average number of individuals	283.1	176.7	283.1	176.7
Minimum number of individuals	101		101	
Maximum number of individuals	1255		1255	
Average LFP	0.351	0.076	0.381	0.075
Minimum LFP	0.138		0.161	
Maximum LFP	0.580		0.580	
Average Fraction 1 vs 2+ children families	0.679	0.060	0.679	0.060
Minimum Fraction 1 vs 2+ children families	0.386		0.386	
MaximumFraction 1 vs 2+ children families	0.838		0.838	

Table 3 - Estimate of the Private Effect of the EITC (without Social Interactions)

Dependent Variable = Labor Force Participation (Annual Earnings > $200)

	Cross Section	Pooled 1993 and 1994	Pooled 1993/94 + FE
(More than 1 child)*(Post 93)		**0.016**	**0.016**
(Private effect of EITC)		**(7.11)****	**(7.07)****
More than 1 child	-0.027	-0.043	
	(9.51)**	(15.51)**	
Post 93		0.019	
		(9.85)**	
Mother's age	0.015	0.017	
	(7.64)**	(9.45)**	
Mother's squared age	-0.000	-0.000	-0.000
	(7.22)**	(8.84)**	(2.27)*
Employment / Population - all sectors	-0.072	-0.095	-0.101
	(5.16)**	(7.44)**	(1.78)
Observations	142391	284782	142391
R-squared	0.0012	0.0029	0.0004

Absolute value of t statistics in parentheses (standard errors in pooled regression clustered by individual)
* significant at 5%; ** significant at 1%

Table 4 - Estimate of the Private, Spillover and Social Effects of the EITC

Dependent Variable = Labor Force Participation (Annual Earnings > $200)

	FE without IV	IV without FE	Baseline: IV + FE
Average LFP	**0.996**	**0.753**	**0.651**
(Spillover effect)	**(472.79)****	**(8.47)****	**(4.07)****
(More than 1 child)*(Post 93)	**0.015**	**0.015**	**0.016**
(Private effect of EITC)	**(6.59)****	**(3.83)****	**(6.59)****
More than 1 child		-0.041	
		(14.24)**	
Post 1993		-0.004	
		(1.09)	
Mother's age		0.015	
		(9.38)**	
Mother's squared age	-0.000	-0.000	-0.000
	(2.20)*	(8.70)**	(2.23)*
Employment / Population - all sectors	-0.104	-0.026	-0.103
	(1.72)	(2.67)**	(1.70)
# Observations	142391	284782	142391
# zip codes	503	503	503
R-squared	0.0046	0.0249	0.0041
First Stage: Dependent Variable = Average Labor Force Participation in the zip code			
Average ((More than 1 child)*(Post 93))			**0.046**
(Social impact of the EITC)			**(2.29)***
R-squared			0.0015
R-squared excluded instrument			0.010

Absolute value of t statistics in parentheses (standard errors allowing for clustering by zip code)

* significant at 5%; ** significant at 1%

Note: Other covariates in first stage (not reported) include all individual variables in second stage

Table 5 - Estimate of the Private, Spillover and Social Effects of the EITC - Robustness to sample definition

Second Stage: Dependent Variable = Labor Force Participation (Annual Earnings > $200)

	zip size > 25	zip size > 50	zip size > 100	zip size > 100	zip size > 100
	Base sample	**Base sample**	**Base sample**	**Sample B***	**Sample C***
Average LFP	**0.574**	**0.544**	**0.651**	**0.629**	**0.603**
(Spillover effect)	**(3.20)****	**(2.32)***	**(4.07)****	**(1.93)**	**(1.56)**
(More than 1 child)*(Post 93)	**0.017**	**0.016**	**0.016**	**0.012**	**0.012**
(Private effect of EITC)	**(7.70)****	**(7.06)****	**(6.59)****	**(3.55)****	**(3.46)****
Squared age	-0.023	-0.022	-0.019	-0.057	-0.063
	(2.90)**	(2.64)**	(2.23)*	(4.48)**	(4.79)**
Employment / Population - all sectors	-0.111	-0.103	-0.103	-0.091	-0.084
	(2.00)*	(1.80)	(1.70)	(1.33)	(1.21)
# Observations	165139	157762	142391	84853	82146
# zip codes	921	716	503	364	354
R-squared	0.0060	0.0047	0.0041	0.0048	0.0048
First Stage: Dependent Variable = Average Labor Force Participation in the zip code					
Average ((More than 1 child)*(Post 93))	**0.041**	**0.036**	**0.046**	**0.034**	**0.031**
(Social impact of the EITC)	**(2.49)***	**(2.01)***	**(2.29)***	**(1.21)**	**(1.08)**
R-squared	0.0066	0.0057	0.0102	0.0052	0.0045
R-squared excluded instrument	0.0060	0.006	0.010	0.004	0.003

Absolute value of t statistics in parentheses (standard errors allowing for clustering by zip code)

* significant at 5%; ** significant at 1%

Note: Other covariates in first stage (not reported) include all individual variables in second stage

Sample B = Base sample restricted to individuals on welfare at least twice between 1992 and 1995

Sample C = Base sample restricted to individuals on welfare at least twice between 1992 and 1995, at most one year apart

Table 6 - Reduced Form estimates, assuming Social Interaction equilibrium

Reduced Form Equation: Dependent Variable = Labor Force Participation (Annual Earnings > $200)

	zip size > 25	zip size > 50	zip size > 100	zip size > 100	zip size > 100
	Base sample	**Base sample**	**Base sample**	**Sample B***	**Sample C***
Average ((More than 1 child)*(Post 93)) (π_1)	**0.02332**	**0.01944**	**0.03020**	**0.02122**	**0.01882**
Implied Spillover effect ($\beta\gamma / (1-\beta)$)	**(1.42)**	**(1.09)**	**(1.48)**	**(0.75)**	**(0.65)**
(More than 1 child)*(Post 93) (π_2)	**0.01671**	**0.01564**	**0.01554**	**0.01196**	**0.01194**
(Private effect of EITC)	**(7.70)****	**(7.07)****	**(6.60)****	**(3.56)****	**(3.47)****
Squared age	-0.00048	-0.00045	-0.00040	-0.00121	-0.00132
	(2.95)**	(2.72)**	(2.28)*	(4.77)**	(5.07)**
Employment / Population - all sectors	-0.10632	-0.09960	-0.09969	-0.06678	-0.06056
	(1.81)	(1.66)	(1.56)	(0.96)	(0.87)
# Observations	165139	157762	142391	84853	82146
# zip codes	921	716	503	364	354
R-squared	0.0005	0.0004	0.0004	0.0005	0.0005
Social interaction estimation, from Reduced Form					
Indirect estimate of $\beta=(\pi_1/(\pi_1+\pi_2))$	**0.5826**	**0.5542**	**0.6603**	**0.6395**	**0.6118**
(Social Interaction coefficient)	**(3.28)****	**(2.38)***	**(4.18)****	**(2.00)***	**(1.60)**

Absolute value of t statistics in parentheses (standard errors allowing for clustering by zip code)

* significant at 5%; ** significant at 1%

Sample B = Base sample restricted to individuals on welfare at least twice between 1992 and 1995

Sample C = Base sample restricted to individuals on welfare at least twice between 1992 and 1995, at most one year apart

Table 7 - Models with (county*time) fixed effect - 2SLS estimates

Second Stage: Dependent Variable = Labor Force Participation (Annual Earnings > $200)

	zip size > 25	**zip size > 50**	**zip size > 100**
	Base sample	**Base sample**	**Base sample**
Average LFP	**0.518**	**0.467**	**0.606**
(Spillover effect)	**(2.25)***	**(1.46)**	**(3.11)****
(More than 1 child)*(Post 93)	**0.017**	**0.016**	**0.016**
(Private effect of EITC)	**(7.69)****	**(7.06)****	**(6.60)****
Squared age	-0.000	-0.000	-0.000
	(2.88)**	(2.60)**	(2.21)*
# Observations	165139	157762	142391
# zip codes	921	716	503
R-squared	0.0058	0.0045	0.0040

First Stage: Dependent Variable = Average Labor Force Participation in the zip code

Average ((More than 1 child)*(Post 93))	**0.038**	**0.032**	**0.045**
(Social impact of the EITC)	**(2.19)***	**(1.73)**	**(2.18)***
R-squared	0.0888	0.0962	0.1154
R-squared excluded instrument	0.005	0.004	0.009

Absolute value of t statistics in parentheses (standard errors allowing for clustering by zip code)
* significant at 5%; ** significant at 1%

Table 8 - Models with (county*time) fixed effect - Reduced Form estimates

Reduced Form Equation: Dependent Variable = Labor Force Participation (Annual Earnings > $200)

	zip size > 25	**zip size > 50**	**zip size > 100**
	Base sample	**Base sample**	**Base sample**
Average ((More than 1 child)*(Post 93)) (π_1)	**0.01954**	**0.01513**	**0.02705**
Implied Spillover effect ($\beta\gamma / (1-\beta)$)	**(1.12)**	**(0.80)**	**(1.30)**
(More than 1 child)*(Post 93) (π_2)	**0.01670**	**0.01564**	**0.01557**
(Private effect of EITC)	**(7.69)****	**(7.06)****	**(6.61)****
Squared age	-0.00047	-0.00044	-0.00039
	(2.92)**	(2.68)**	(2.24)*
# Observations	165139	157762	142391
# zip codes	921	716	503
R-squared	0.0011	0.0010	0.0010
Social interaction estimation, from Reduced Form			
Indirect estimate of $\beta=(\pi_1/(\pi_1+\pi_2))$	**0.5392**	**0.4917**	**0.6347**
(Social Interaction coefficient)	**(2.38)***	**(1.54)**	**(3.42)****

Absolute value of t statistics in parentheses (standard errors allowing for clustering by zip code)
* significant at 5%; ** significant at 1%

Table 9 – Summary Statistics – Chapter 3

Variable	Mean	Standard Deviation
FG family	0.84	0.37
# Kids under 1	0.28	0.49
# Kids under 5	0.91	0.90
Age oldest person	32.62	8.76
Oldest person in case is of White race	0.31	0.46
Oldest person in case is of Black race	0.23	0.42
Oldest person in case is of Hispanic Origin	0.32	0.47
Real Maximum EITC benefit (1999 US$)	2906.1	807.9
Whether filed a tax return	0.43	0.50
Whether claimed the EITC	0.40	0.49
Whether worked (EDD earnings > 0)	0.53	0.50
Whether claimed, conditional on filed	0.92	
Whether filed, conditional on worked	0.70	
Whether claimed, conditional on worked (take-up)	64.4%	
zip code population 1990	42101.2	19212.7
zip code population 2000	46072.5	21282.4
# of sampled individuals / zip code	990.9	685.6
Number of TPS / zip code / (1000) population in 2000	0.07	0.47
Number of TPS / zip code	2.16	1.41
County Average Employment / Population ratio	0.63	0.10

Table 10 - Evolution of Tax Preparation Services over time, as a function of its EITC potential

	A) Real potential EITC benefit based on earnings distribution, assuming all families w/2 children				B) Real potential EITC benefit based on earnings distribution, assuming all families w/1 child				C) Real potential EITC benefit based on poor families w/children, assuming all families w/2 children				D) Real potential EITC benefit based on poor families w/children, assuming all families w/1 child			
Dep. Variable	Number of TPS		TPS > 0		Number of TPS		TPS > 0		Number of TPS		TPS > 0		Number of TPS		TPS > 0	
EITC measure (1999 US$)	-0.015 (1.80)		-0.011 (3.06)**		-0.045 (2.16)*		-0.029 (3.33)**		-0.052 (3.15)**		-0.024 (3.48)**		-0.143 (3.25)**		-0.065 (3.52)**	
EITC*small zip		-0.058 (1.75)		0.001 (0.08)		-0.145 (1.73)		0.012 (0.34)		-0.056 (0.75)		0.047 (1.52)		-0.058 (1.75)		0.001 (0.08)
EITC*big zip		-0.017 (2.01)*		-0.010 (2.87)**		-0.049 (2.32)*		-0.027 (3.10)**		-0.052 (3.15)**		-0.024 (3.42)**		-0.017 (2.01)*		-0.010 (2.87)**
year==1992	-0.035 (0.72)	-0.033 (0.68)	-0.012 (0.59)	-0.012 (0.61)	-0.029 (0.60)	-0.025 (0.52)	-0.009 (0.42)	-0.010 (0.49)	-0.035 (0.72)	-0.035 (0.72)	-0.013 (0.63)	-0.013 (0.66)	-0.031 (0.63)	-0.033 (0.68)	-0.011 (0.54)	-0.012 (0.61)
year==1993	0.235 (4.83)**	0.238 (4.89)**	0.104 (5.09)**	0.103 (5.05)**	0.244 (4.98)**	0.250 (5.08)**	0.109 (5.33)**	0.107 (5.18)**	0.235 (4.85)**	0.235 (4.85)**	0.102 (5.04)**	0.101 (4.99)**	0.243 (4.99)**	0.238 (4.89)**	0.106 (5.20)**	0.103 (5.05)**
year==1994	0.888 (17.73)**	0.900 (17.68)**	0.293 (13.98)**	0.289 (13.59)**	0.906 (17.41)**	0.923 (17.15)**	0.303 (13.93)**	0.296 (13.16)**	0.900 (18.11)**	0.900 (17.95)**	0.293 (14.08)**	0.286 (13.65)**	0.925 (17.85)**	0.900 (17.68)**	0.304 (14.02)**	0.289 (13.59)**
year==1995	0.970 (18.66)**	0.989 (18.37)**	0.326 (15.00)**	0.321 (14.26)**	0.985 (18.42)**	1.006 (17.97)**	0.334 (14.91)**	0.325 (13.90)**	0.986 (19.35)**	0.987 (19.02)**	0.325 (15.25)**	0.315 (14.55)**	0.998 (19.19)**	0.989 (18.37)**	0.330 (15.17)**	0.321 (14.26)**
year==1996	0.948 (17.40)**	0.973 (16.91)**	0.337 (14.80)**	0.330 (13.74)**	0.957 (17.57)**	0.979 (17.07)**	0.340 (14.92)**	0.331 (13.78)**	0.963 (18.50)**	0.964 (18.05)**	0.333 (15.29)**	0.322 (14.40)**	0.966 (18.52)**	0.973 (16.91)**	0.334 (15.29)**	0.330 (13.74)**
year==1997	1.058 (19.19)**	1.084 (18.54)**	0.359 (15.59)**	0.352 (14.41)**	1.069 (19.29)**	1.093 (18.61)**	0.363 (15.68)**	0.353 (14.39)**	1.072 (20.54)**	1.072 (20.04)**	0.354 (16.22)**	0.342 (15.30)**	1.074 (20.56)**	1.084 (18.54)**	0.355 (16.23)**	0.352 (14.41)**
year==1998	1.186 (21.22)**	1.214 (20.38)**	0.384 (16.42)**	0.376 (15.10)**	1.199 (21.20)**	1.225 (20.28)**	0.388 (16.43)**	0.378 (14.96)**	1.197 (22.92)**	1.198 (22.34)**	0.376 (17.23)**	0.365 (16.27)**	1.201 (22.91)**	1.214 (20.38)**	0.377 (17.22)**	0.376 (15.10)**
year==1999	1.145 (20.45)**	1.173 (19.64)**	0.373 (15.92)**	0.365 (14.62)**	1.159 (20.36)**	1.186 (19.46)**	0.378 (15.90)**	0.367 (14.42)**	1.155 (22.17)**	1.156 (21.64)**	0.365 (16.74)**	0.353 (15.82)**	1.158 (22.18)**	1.173 (19.64)**	0.365 (16.74)**	0.365 (14.62)**
Constant	0.758 (21.12)**	0.768 (20.93)**	0.491 (32.75)**	0.488 (31.82)**	0.794 (18.57)**	0.817 (17.52)**	0.513 (28.71)**	0.504 (25.85)**	0.777 (21.38)**	0.777 (20.95)**	0.495 (32.59)**	0.488 (31.47)**	0.840 (18.14)**	0.768 (20.93)**	0.524 (27.04)**	0.488 (31.82)**
Observations	5688	5688	5688	5688	5688	5688	5688	5688	5688	5688	5688	5688	5688	5688	5688	5688
# of zip codes	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632	632
R-squared	0.23	0.24	0.15	0.15	0.24	0.24	0.15	0.15	0.24	0.24	0.15	0.15	0.24	0.24	0.15	0.15

Absolute value of t statistics in parentheses (* significant at 5%; ** significant at 1%). All specifications with zip code fixed effects and year effects (ommited year is 1991).

Notes:

1) "Small zip" corresponds to zip codes with less than 22,090 individuals, based on 100% population counts from Census 1990.

2) EITC measures are constructed as following:

A) Obtain number of families for each earnings range (0-5000, 5000-10000, …) from Census 1990 and assign the maximum 2-children real EITC benefit for the family in the midpoint. These benefits are added over all income ranges eligible for EITC

B) Similar to A but using the 1-child max EITC benefits

C) Obtain number of families with children and earnings below poverty level from Census 1990 and assign the 2-children maximum real EITC benefit for the family in the flat range.

D) Similar to C but using the 1-child max EITC benefits

Table 11 - Effect of Tax Preparation Services on EITC participation and Tax Filing status
Zip codes with at least 50 sampled families

	Dependent variable			
	Whether claimed the EITC		Whether filed a tax return	
	OLS	zip FE	OLS	zip FE
TPS / Population 2000	4.648	16.110	5.004	15.172
	(5.04)**	(5.50)**	(5.49)**	(5.12)**
Maximum EITC benefit (thousand 1999 US$)	0.0948	0.0495	0.0906	0.0425
	(56.24)**	(88.98)**	(52.90)**	(75.45)**
County Employment to population ratio	0.088	3.600	0.130	3.892
	(3.74)**	(174.93)**	(5.11)**	(186.99)**
# Kids under age 1	-0.055	-0.054	-0.057	-0.055
	(47.14)**	(60.35)**	(47.28)**	(61.16)**
# Kids under age 5	-0.060	-0.047	-0.064	-0.049
	(65.15)**	(86.60)**	(70.77)**	(90.49)**
age oldest person	-0.003	0.003	-0.005	0.001
	(4.55)**	(11.81)**	(8.02)**	(4.84)**
(age oldest person)^2 (*1000)	-0.036	-0.117	-0.0131	-0.0982
	(4.27)**	(30.46)**	(1.47)	(25.39)**
FG case (0 if UP case)	-0.272	-0.273	-0.274	-0.277
	(60.94)**	(259.89)**	(63.62)**	(260.91)**
White	-0.009	-0.025	0.012	-0.007
	(1.26)	(19.29)**	(1.62)	(5.32)**
Black	0.054	0.066	0.059	0.069
	(6.41)**	(46.03)**	(7.00)**	(47.83)**
Hispanic	0.066	0.061	0.074	0.069
	(8.46)**	(47.71)**	(9.84)**	(53.49)**
Constant	0.476	-1.720	0.544	-1.806
	(26.27)**	(124.84)**	(28.09)**	(129.69)**
Observations	1,741,442	1,741,442	1741442	1741442
R-squared	0.07	0.08	0.07	0.08
Number of zip codes		664		664

Absolute value of t statistics in parentheses.
* significant at 5%; ** significant at 1%

References

Aaronson, D., "Sibling estimates of neighborhood effects", in Brooks-Gunn, J., G.J. Duncan, & L. Aber, Eds. Neighborhood Poverty: Policy Implications in Studying Neighborhoods, Vol II. (pp. 80-93). New York: Russell Sage Foundation.

Aizer, A. and Currie, J., "Networks or Neighborhoods? Correlations in the Use of Publicly-Funded Maternity Care in California", *NBER Working Paper* No.w9209, September 2002.

Aizer, Anna, "Covering Kids: Efforts to Increase the Health Insurance Coverage of Poor Children", mimeo UCLA, August 2001.

Aizer, Anna, "Got Health? Advertising, Medicaid and Child Health", mimeo Princeton University, March 5, 2003.

Arrow, K. and R. Borzekowski, "Limited Network Connections and the Distribution of Wages", mimeo Stanford University, 2000.

C. Avery and P. Zemsky, "Multidimensional uncertainty and herd behavior in financial markets", *American Economic Review* 88 (1998), 724—748.

A. V. Banerjee, "A simple model of herd behavior", *Quarterly Journal of Econonomics*, 107 (1992), 797—817.

Bertrand, Luttmer and Mullainathan, "Network Effects and Welfare Cultures", *Quarterly Journal of Economics* 115, August 2000, pages 1019-1056.

Berube, A., Kim, A., Forman, B., and Burns, M., "The Price of Paying Taxes: How tax Preparation and Refund Loan Fees Erode the Benefits of the EITC", The Brookings Institution, Progressive Policy Institute, Survey Series, May 2002.

Besley and Coate, "Understanding Welfare Stigma: Taxpayer Resentment and Statistical Discrimination", *Journal of Public Economics*, 48(2), July 1992, pages 165-83.

S. Bikhchandani, D. Hirshleifer, and I. Welch, "A theory of fads, fashion, custom, and cultural change in informational cascades", *Journal of Political Economy* 100, 1992, 992-1026.

S. Bikhchandani, D. Hirshleifer, and I. Welch, "", *Journal of Economic Perspectives*, 1998.

Blank and Ruggles, "When Do Women Use Aid to Families with Dependent Children and Food Stamps? The Dynamics of Eligibility versus Participation", *Journal of Human Resources*, 31(1), pages 57-89, 1996.

Boorman, S.A., "A Combinatorial Optimization Model for Transmission of Job Information through Contact Networks", *Bell Journal of Economics* 6, 1975, 216-249.

Borjas, George, "Ethnic Capital and Intergenerational Mobility", *Quarterly Journal of Economics*, 1991, 108, pages 123-150.

Borjas, George, "Ethnicity, Neighborhoods, and Human-Capital Externalities", *American Economic Review*, June 1995, pages 365-390.

Brock, William A. and Durlauf, Steven N., "Interactions-Based Models", *Handbook of Econometrics*, Volume 5, Edited by J.J. Heckman and E. Leamer, 2001.

Burtless, Gary, "The Case for Randomized Field Trials in Economics and Policy Research", *The Journal of Economic Perspectives*, Volume 9, Issue 2 (Spring, 1995), pages 63-84.

Case, A. and L. Katz, "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths", *NBER Working Paper* No. 9375, May 1991.

Calvó-Armengol, Antoni and Jackson, Matthew O. (2000), "Job Contact Networks", forthcoming in *Journal of Economic Theory*.

Calvó-Armengol, Antoni and Jackson, Matthew O., "The Effects of Social Networks on Unemployment and Inequality", mimeo Universitat Autonoma de Barcelona, September 2003.

Chari, V.V. and Kehoe, P.J., "Financial Crises As Herds: Overturning The Critiques", *NBER Working Paper* 9658, April 2003.

Currie, Janet and Jeffrey Grogger. "Explaining Recent Declines in Food Stamp Program Participation", in *Brookings Papers on Urban Affairs*, William Gale and Janet Rothenberg-Pack (eds), 2001.

Cutler, D and E. Glaeser, "Are Ghettos Good or Bad", *Quarterly Journal of Economics* 112, August 1997, pages 827-872.

Daponte, Sanders and Taylor, "Why Do Low-Income Households Not Use Food Stamps? Evidence from an Experiment", *Journal of Human Resources*, 34(3), Summer 1999, pages 612-28.

Dickert, Stacy, Scott Houser and John Karl Scholz, 1995, "The Earned Income Tax Credit and Transfer Programs: A Study of Labor Market and Program Participation, Tax Policy and the Economy", James M. Poterba (ed.), National Bureau of Economic Research and the MIT Press 9, 1-50.

Duflo, E. and Saez, E., "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment", *The Quarterly Journal of Economics*, Vol. 118, Issue 3, August 2003.

Duncan, Greg and Raudenbush, Stephen, "Neighborhoods and Adolescent Development: How Can we Determine the Links?" working paper, Northwestern University.

Eissa, N. and Hoynes, H., "The Earned Income Tax Credit and the Labor Supply of Married Couples", *NBER Working Paper* 6856, December 1998.

Eissa, Nada and Jeffrey B. Liebman, "Labor Supply Response to the Earned Income Tax Credit", *Quarterly Journal of Economics*, May 1996, 605-637.

Ellison, Glen and Fudenberg, Drew, "Rules of Thumb for Social Learning", *The Journal of Political Economy*, Vol 101, No 4, August 1993, 612-643.

Ellwood, David T., 2000, "The Impact of the Earned Income Tax Credit and Social Policy Reforms on Work, Marriage, and Living Arraignments", mimeo, Kennedy School of Government.

Evans, Oates and Schwab, "Measuring Peer Group Effects: A Study of Teenage Behavior", *Journal of Political Economy*, 1992, vol. 100, no. 5.

General Accounting Office, "Food Stamps: Examination of Program Data and Analysis of Nonparticipation", Washington, D.C., GAO, 1988.

Glaeser, Sacerdote and Scheinkman, "Crime and Social Interactions", *Quarterly Journal of Economics*, May 1996, pages 507-548.

Glaeser, Sacerdote and Scheinkman, "The Social Multiplier", *NBER Working Paper* No.w9153, September 2002.

L. R. Glosten and P. R. Milgrom, "Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders", *Journal of Financial Economics* 14, 1985, 71-100.

Griliches, Zvi, "Hybrid Corn: An Exploration in the Economics of Technological Change", *Econometrica* 25, October 1957, 501-522.

Grogger, Jeffrey, "Welfare Transitions in the 1990s: The Economy, Welfare Policy, and the EITC", *NBER Working Paper* w9472, February 2003.

Grogger, Jeffrey, "The Effects of Time Limits, the EITC, and Other Policy Changes on Welfare Use, Work, and Income among Female-Headed Families", *The Review of Economics and Statistics*, Vol. 85, Issue 2, May 2003.

Heckman James and Smith, Jeffrey, "Assessing the Case for Social Experiments", *The Journal of Economic Perspectives*, Volume 9, Issue 2 (Spring, 1995), pages 85-110.

Hill, C., Hotz, V. Joseph, Mullin, C., and Scholz, J. K., "EITC Eligibility, Participation and Compliance Rates for AFDC Households: Evidence from the California Caseload", Final report to the State of California, University of California, Los Angeles, April 1999.

Hill, Daniel, "An Endogenously-Switching Ordered-Response Model of Information, Eligibility and Participation in SSI", *Review of Economics and Statistics*, 72(2), May 1990, pages 368-71.

Hotz, Joseph V. and Scholz, John Karl, "The Earned Income Tax Credit", in Means-Tested Transfer Programs in the United States, edited by Robert Moffitt, The University of Chicago Press, 2003.

Hotz, V. Joseph, Charles H. Mullin and John Karl Scholz, "The Earned Income Tax Credit and Labor Market Participation of Families on Welfare", mimeo, UCLA, Vanderbilt and University of Wisconsin Madison, March 2002.

Caroline Hoxby, "Peer Effects in the Classroom: Learning from Gender and Race Variation", *NBER Working Paper* No. 7867, August 2000.

Katz, Lawrence, Kling, Jeffrey and Liebman, Jeffrey, "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment", *NBER Working paper* 7973, October 2000.

Kooreman, P. and Soetevent, A., "A Discrete Choice Model with Social Interactions; an Analysis of High School Teen Behavior", mimeo, University of Groningen, February 2002.

Kyriazidou, Ekaterini, "Estimation of A Panel Data Sample Selection Model," *Econometrica*, Vol. 65, No 6, pp. 1335-1364 (1997).

I. H. Lee, "On the convergence of informational cascades", *Journal of Economic Theory* 61, 1993, 395—411.

Ludwig, Jens, Duncan, Greg, and Pinkston, Joshua, "Neighborhood Effects on Economic Self-Sufficiency: Evidence from a Randomized Housing-Mobility Experiment", mimeo Georgetown Public Policy Institute, January 31, 2000.

Mansfield, Edwin, "The Economics of Technological Change", New York: Norton, 1968.

Charles Manski, "Identification of Endogenous Social Effects: The Reflection Problem", *Review of Economic Studies,* 60, 1993, pages 531-542.

Charles Manski, "Economic Analysis of Social Interactions", *Journal of Economic Perspectives*, forthcoming.

Marmaros, David and Sacerdote, Bruce, "Peer and social networks in job search", *European Economic Review* 46 (2002), 870-879.

Meyer, B. and Rosenbaum, D., "Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers", *Quarterly Journal of Economics*, August 2001; 116(3): 1063-1114.

Robert Moffitt, "An Economic Model of Welfare Stigma", *American Economic Review*, December 1983.

Robert Moffitt, "Policy Interventions, Low-Level Equilibria and Social Interactions", Forthcoming in *Social Dynamics*, eds. Steven Durlauf and Peyton Young, MIT Press.

Montgomery, J., "Social Networks and Labor Market Outcomes: Towards an Economic Analysis", *American Economic Review* 81, 1991, 1408-1418.

Moretti, E., "Estimating the Social Return to Higher Education: Evidence From Longitudinal and Cross-Sectional Data", forthcoming in *Journal of Econometrics*.

Rank and Hirschl, "The Link between Population Density and Welfare Participation", *Demography*; 30(4), November 1993, pages 607-22.

Ross P., Katherin, "Who Knows about the Earned Income Tax Credit?", New Federalism National Survey of America's Families, Series B, No. B-27, January 2001.

Sacerdote, B., "Peer Effects with Random Assignment: Results for Dartmouth Roommates", *Quarterly Journal of Economics*. May 2001; 116(2): 681-704.

Scholz, J.K., "The Earned Income Tax Credit: Participation, Compliance And Anti-Poverty Effectiveness", *National Tax Journal*, March 1994, pages 63-87.

Strauss, Robert, "Information and Participation in a Public Transfer Program", *Journal of Public Economics* 8(3), Dec. 1977, pages 385-96.

Topa, G., "Social Interactions, Local Spillovers and Unemployment", *Review of Economic Studies 68*, 2001, 261-295.