### Table 1
Parameters of the 11 Negative Income Tax Programs

<table>
<thead>
<tr>
<th>Program Number</th>
<th>G ($)</th>
<th>τ</th>
<th>Declining Tax Rate</th>
<th>Break-even Income ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3,800</td>
<td>.5</td>
<td>No</td>
<td>7,600</td>
</tr>
<tr>
<td>2</td>
<td>3,800</td>
<td>.7</td>
<td>No</td>
<td>5,429</td>
</tr>
<tr>
<td>3</td>
<td>3,800</td>
<td>.7</td>
<td>Yes</td>
<td>7,367</td>
</tr>
<tr>
<td>4</td>
<td>3,800</td>
<td>.8</td>
<td>Yes</td>
<td>5,802</td>
</tr>
<tr>
<td>5</td>
<td>4,800</td>
<td>.5</td>
<td>No</td>
<td>9,600</td>
</tr>
<tr>
<td>6</td>
<td>4,800</td>
<td>.7</td>
<td>No</td>
<td>6,857</td>
</tr>
<tr>
<td>7</td>
<td>4,800</td>
<td>.7</td>
<td>Yes</td>
<td>12,000</td>
</tr>
<tr>
<td>8</td>
<td>4,800</td>
<td>.8</td>
<td>Yes</td>
<td>8,000</td>
</tr>
<tr>
<td>9</td>
<td>5,600</td>
<td>.5</td>
<td>No</td>
<td>11,200</td>
</tr>
<tr>
<td>10</td>
<td>5,600</td>
<td>.7</td>
<td>No</td>
<td>8,000</td>
</tr>
<tr>
<td>11</td>
<td>5,600</td>
<td>.8</td>
<td>Yes</td>
<td>10,360</td>
</tr>
</tbody>
</table>

Source: Ashenfelter and Plant (1990), p. 403

Source: Ashenfelter and Plant (1990)
FIGURE 2. PROPORTION WITH POSITIVE EARNINGS FOR NONWINNERS, WINNERS, AND BIG WINNERS

Note: Solid line = nonwinners; dashed line = winners; dotted line = big winners.

On average the individuals in our basic sample won yearly prizes of $26,000 (averaged over the $55,000 for winners and zero for nonwinners). Typically they won 10 years prior to completing our survey in 1996, implying they are on average halfway through their 20 years of lottery payments when they responded in 1996. We asked all individuals how many tickets they bought in a typical week in the year they won the lottery. As expected, the number of tickets bought is considerably higher for winners than for nonwinners. On average, the individuals in our basic sample are 50 years old at the time of winning, which, for the average person was in 1986; 35 percent of the sample was over 55 and 15 percent was over 65 years old at the time of winning; 63 percent of the sample was male. The average number of years of schooling, calculated as years of high school plus years of college plus 8, is equal to 13.7; 64 percent claimed at least one year of college.

We observe, for each individual in the basic sample, Social Security earnings for six years preceding the time of winning the lottery, for the year they won (year zero), and for six years following winning. Average earnings, in terms of 1986 dollars, rise over the pre-winning period from $13,930 to $16,330, and then decline back to $13,290 over the post-winning period. For those with positive Social Security earnings, average earnings rise over the entire 13-year period from $20,180 to $24,300. Participation rates, as measured by positive Social Security earnings, gradually decline over the 13 years, starting at around 70 percent before going down to 56 percent. Figures 1 and 2 present graphs for average earnings and the proportion of individuals with positive earnings for the three groups, nonwinners, winners, and big winners. One can see a modest decline in earnings and proportion of individuals with positive earnings for the full winner sample compared to the nonwinners after winning the lottery, and a sharp and much larger decline for big winners at the time of winning. A simple difference-in-differences type estimate of the marginal propensity to earn out of unearned income (mpe) can be based on the ratio of the difference in the average change in earnings before and after winning the lottery for two groups and the difference in the average prize for the same two groups.

For the winners, the difference in average earnings over the six post-lottery years and the six pre-lottery years is -$1,877 and for the nonwinners the average change is $448. Given a difference in average prize of $55,000 for the winner/nonwinners comparison, the estimated mpe is \((-1,877 - 448) / (55,000 - 0) = -0.042\) (SE 0.016). For the big-winners/small-winners comparison, this estimate is -0.059 (SE 0.018). In Section IV we report estimates for this quantity using more sophisticated analyses.

On average the value of all cars was $18,200. For housing the average value was $166,300, with an average mortgage of $44,200.$12 We aggregated the responses to financial wealth into two categories. The first concerns retirement

Note: Solid line = nonwinners; dashed line = winners; dotted line = big winners.

Source: Imbens et al. (2001), p. 783
Figure II

Source: Federal Govt

marginal tax rate

Taxable Income (1985$)

high-income

75th percentile

90th percentile

28

33

33

42

45

50

15

28

59700

124000

Source: Federal Govt
## Table IIa
### Marginal Tax Rate

<table>
<thead>
<tr>
<th>Group</th>
<th>Before TRA86</th>
<th>After TRA86</th>
<th>Change</th>
<th>Relative Change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>0.521 (.002)</td>
<td>0.382 (.001)</td>
<td>-0.139</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75th Percentile</td>
<td>0.365 (.001)</td>
<td>0.324 (.001)</td>
<td>-0.041</td>
<td>-0.098 (.002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90th Percentile</td>
<td>0.430 (.001)</td>
<td>0.360 (.001)</td>
<td>-0.07</td>
<td>-0.069 (.002)</td>
</tr>
</tbody>
</table>

The marginal tax rate is calculated using family wage and salary, self-employment, interest, dividend, farm and social-security income. I assume all couples file jointly, and that all itemize their deductions. Itemized deductions and capital gains are imputed using Statistics of Income data. These figures include the secondary earner deduction, as well as social security taxes. Standard errors are in parentheses. Before TRA86 is tax years 1983-1985; After TRA86 is tax years 1989-1991.

Source: Eissa 1995
### Table III
Differences-in-Differences Estimates
CPS Married Women Before and After TRA86

**A: Labor Force Participation**

<table>
<thead>
<tr>
<th>Group</th>
<th>Before TRA86</th>
<th>After TRA86</th>
<th>Change</th>
<th>Difference-in-Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>0.464 (.018)</td>
<td>0.554 (.018)</td>
<td>0.090 (.025)</td>
<td>{19.5%}</td>
</tr>
<tr>
<td></td>
<td>[756]</td>
<td>[718]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>75th Percentile</td>
<td>0.687 (.010)</td>
<td>0.740 (.010)</td>
<td>0.053 (.010)</td>
<td>0.037 (.028)</td>
</tr>
<tr>
<td></td>
<td>[3799]</td>
<td>[3613]</td>
<td></td>
<td>{12.3%}</td>
</tr>
<tr>
<td>90th Percentile</td>
<td>0.611 (.010)</td>
<td>0.656 (.010)</td>
<td>0.045 (.010)</td>
<td>0.045 (.028)</td>
</tr>
<tr>
<td></td>
<td>[3765]</td>
<td>[3584]</td>
<td></td>
<td>{13%}</td>
</tr>
</tbody>
</table>

Source: Eissa 1995
B: Hours Conditional on Employment

<table>
<thead>
<tr>
<th>Group</th>
<th>Before TRA86</th>
<th>After TRA86</th>
<th>Change</th>
<th>Difference-in-Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>1283.0 (46.3)</td>
<td>1446.3 (41.1)</td>
<td>163.3 (61.5)</td>
<td>{12.7%}</td>
</tr>
<tr>
<td></td>
<td>[351]</td>
<td>[398]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>75th Percentile</td>
<td>1504.1 (14.3)</td>
<td>1558.9 (13.9)</td>
<td>54.8 (20.0)</td>
<td>108.6 (65.1)</td>
</tr>
<tr>
<td></td>
<td>[2610]</td>
<td>[2676]</td>
<td></td>
<td>{9.4%}</td>
</tr>
<tr>
<td>90th Percentile</td>
<td>1434.1 (16.4)</td>
<td>1530.1 (15.9)</td>
<td>96.0 (22.8)</td>
<td>67.3 (64.8)</td>
</tr>
<tr>
<td></td>
<td>[2303]</td>
<td>[2348]</td>
<td></td>
<td>{6.2%}</td>
</tr>
</tbody>
</table>

Each cell contains the mean for that group, along with standard errors in (), number of observations in [], and % increase in {}. Means are unweighted.

Source: Eissa 1995
Figure 10
Fraction of Married Women with Positive Annual Earnings by Income Group in March CPS

Notes: Groups are based on other household income (husband’s earnings plus asset income) as described in Eissa (1995). Group 1 <= 75\textsuperscript{th} percentile. Group 75 is >75\textsuperscript{th} percentile and <= 80\textsuperscript{th} percentile. Group 80 is >80\textsuperscript{th} and <= 90\textsuperscript{th}. Group 90 is >90\textsuperscript{th} and <= 95\textsuperscript{th}. Group 95 is >95\textsuperscript{th} and <= 99\textsuperscript{th}. Group 99 is >99\textsuperscript{th}.

Source: Lieberman and Saez (2006)
and control groups. Unfortunately, these data have some critical limitations relative to the administratively based Income Assistance data. Most importantly, they are only available for 52 months after random assignment. Since some program group members were still receiving subsidy payments as late as month 52, this time window is too short to assess the long-run effects of the program. Indeed, looking at Figure 1a, there is still an impact on IA participation in month 52 that does not fully dissipate until month 69. Second, because of nonresponses and refusals, labor market information is only available for 85% of the experimental sample (4,757 people). Third, there appear to be relatively large recall errors and seam biases in the earnings and wage data. Nevertheless, the labor market outcomes provide a valuable complement to the administratively based welfare participation data.

Figure 3 shows the average monthly employment rates of the program and control groups, along with the associated experimental impacts. After random assignment the employment rate of the control group shows a steady

![Figure 3. Monthly employment rates.](source: Card and Hyslop, 2005, p. 1734)

18The distribution of response patterns to the 18-, 36-, and 54-month surveys is fairly similar for the program and control groups (chi-squared statistic = 11.4 with 7 degrees of freedom, p-value = 0.12). However, a slightly larger fraction of the program group have complete labor market data for 52 months—85.4% versus 84.0% for the controls. Moreover, the difference in mean IA participation between the treatment and control groups in month 52 is a little different in the overall sample (2.5%) than in the subset with complete labor market histories (3.3%).

19Each of the three post-random-assignment surveys asked people about their labor market outcomes in the 18 months since the previous survey. Many people report constant earnings over the recall period, leading to a pattern of measured pay increases that are concentrated at the seams, rather than occurring more smoothly over the recall period.
How Does the Current Assistance Caseload Level Compare with Historical Levels?

Figure 2 provides a long-term historical perspective on the number of families receiving assistance from TANF or its predecessor program, from July 1959 to September 2017. The shaded areas of the figure represent months when the national economy was in recession. Though the health of the national economy has affected the trend in the cash assistance caseload, the long-term trend in receipt of cash assistance does not follow a classic countercyclical pattern. Such a pattern would have the caseload rise during economic slumps, and then fall again during periods of economic growth. Factors other than the health of the economy (demographic trends, policy changes) also have influenced the caseload trend.

The figure shows two periods of sustained caseload increases: the period from the mid-1960s to the mid-1970s and a second period from 1988 to 1994. The number of families receiving assistance peaked in March 1994 at 5.1 million families. The assistance caseload fell rapidly in the late 1990s (after the 1996 welfare reform law) before leveling off in 2001. In 2004, the caseload began another decline, albeit at a slower pace than in the late 1990s. During the recent 2007-2009 recession and its aftermath, the caseload began to rise from 1.7 million families in August 2008, peaking in December 2010 at close to 2.0 million families. By September 2018, the assistance caseload had declined to 1.2 million families.

Figure 2. Number of Families Receiving Cash Assistance, July 1959-September 2018

Source: Congressional Research Service (CRS) with data from the U.S. Department of Health and Human Services (HHS).

Notes: Shaded areas denote months when the national economy was in recession. Information represents families receiving cash assistance from Aid to Dependent Children (ADC), Aid to Families with Dependent Children (AFDC), and TANF. For October 1999 through September 2018, includes families receiving assistance from Separate State Programs (SSPs) with expenditures countable toward the TANF maintenance of effort requirement. See Table A-1 for average annual data on families, recipients, adult recipients, and child recipients of ADC, AFDC, and TANF cash assistance for 1961 to 2017.

Table B-5 shows recent trends in the number of cash assistance families by state.
The landscape providing assistance to poor families with children has changed substantially.

![Graph showing changes in welfare programs](image)

Contractions
AFDC/TANF Cash Benefits Per Capita
Food Stamp Expenditures Per Capita
EITC Expenditures Per Capita

Annual Employment Rates for Women
By Marital Status and Presence of Children, 1980-2009

EITC Amount as a Function of Earnings

Source: Federal Govt
FIGURE IV
1986 and 1988 Earned Income Tax Credit

<table>
<thead>
<tr>
<th>TABLE II</th>
<th>LABOR FORCE PARTICIPATION RATES OF UNMARRIED WOMEN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-TRA86 (1)</td>
</tr>
<tr>
<td>A.</td>
<td></td>
</tr>
<tr>
<td>Treatment group:</td>
<td></td>
</tr>
<tr>
<td>With children</td>
<td>0.729 (0.004)</td>
</tr>
<tr>
<td>[20,810]</td>
<td></td>
</tr>
<tr>
<td>Control group:</td>
<td></td>
</tr>
<tr>
<td>Without children</td>
<td>0.952 (0.001)</td>
</tr>
<tr>
<td>[46,287]</td>
<td></td>
</tr>
<tr>
<td>B.</td>
<td></td>
</tr>
<tr>
<td>Treatment group:</td>
<td></td>
</tr>
<tr>
<td>Less than high school, with children</td>
<td>0.479 (0.010)</td>
</tr>
<tr>
<td>[5396]</td>
<td></td>
</tr>
<tr>
<td>Control group 1:</td>
<td></td>
</tr>
<tr>
<td>Less than high school, without children</td>
<td>0.784 (0.010)</td>
</tr>
<tr>
<td>[3958]</td>
<td></td>
</tr>
<tr>
<td>Control group 2:</td>
<td></td>
</tr>
<tr>
<td>Beyond high school, with children</td>
<td>0.911 (0.005)</td>
</tr>
<tr>
<td>[5712]</td>
<td></td>
</tr>
<tr>
<td>C.</td>
<td></td>
</tr>
<tr>
<td>Treatment group:</td>
<td></td>
</tr>
<tr>
<td>High school, with children</td>
<td>0.764 (0.006)</td>
</tr>
<tr>
<td>[9702]</td>
<td></td>
</tr>
<tr>
<td>Control group 1:</td>
<td></td>
</tr>
<tr>
<td>High school, without children</td>
<td>0.945 (0.002)</td>
</tr>
<tr>
<td>[16,527]</td>
<td></td>
</tr>
<tr>
<td>Control group 2:</td>
<td></td>
</tr>
<tr>
<td>Beyond high school, with children</td>
<td>0.911 (0.005)</td>
</tr>
<tr>
<td>[5712]</td>
<td></td>
</tr>
</tbody>
</table>


All Unmarried Females

Figure II
elasticity $e$ would no longer be a pure compensated elasticity, but a mix of the compensated elasticity and the uncompensated elasticity. Four points should be noted.

First, the larger the behavioral elasticity, the more bunching we should expect.

Unsurprisingly, if there are no behavioral responses to marginal tax rates, there

**Figure 1. Bunching Theory**

Notes: Panel A displays the effect on earnings choices of introducing a (small) kink in the budget set by increasing the tax rate $t$ by $dt$ above income level $z^*$. Individual $L$ who chooses $z^*$ before the reform stays at $z^*$ after the reform. Individual $H$ chooses $z^*$ after the reform and was choosing $z^* + dz^*$ before the reform. Panel B depicts the effects of introducing the kink on the earnings density distribution. The pre-reform density is smooth around $z^*$. After the reform, all individuals with income between $z^*$ and $z^* + dz^*$ before the reform, bunch at $z^*$, creating a spike in the density distribution. The density above $z^* + dz^*$ shifts to $z^*$ (so that the resulting density and is no longer smooth at $z^*$).
B. Density Distributions and Bunching

Before tax income $z$

Before reform density

Pre-reform incomes between $z^*$ and $z^* + dz^*$ bunch at $z^*$ after reform

After reform density

Before tax income $z$
indexes earnings to 2008 using the IRS inflation parameters, so that the EITC kinks are perfectly aligned for all years.

Two elements are worth noting in Figure 3. First, there is a clear clustering of tax filers around the first kink point of the EITC. In both panels, the density is maximum exactly at the first kink point. The fact that the location of the first kink point differs between EITC recipients with one child, versus those with two or more children, constitutes strong evidence that the clustering is driven by behavioral responses to the EITC as predicted by the standard model. Second, however, we cannot discern any

**Figure 3. Earnings Density Distributions and the EITC**

*Notes:* The figure displays the histogram of earnings (by $500 bins) for tax filers with one dependent child (panel A) and tax filers with two or more dependent children (panel B). The histogram includes all years 1995–2004 and inflates earnings to 2008 dollars using the IRS inflation parameters (so that the EITC kinks are aligned for all years). Earnings are defined as wages and salaries plus self-employment income (net of one-half of the self-employed payroll tax). The EITC schedule is depicted in dashed line and the three kinks are depicted with vertical lines. Panel A is based on 57,692 observations (representing 116 million tax returns), and panel B on 67,038 observations (representing 115 million returns).
indexes earnings to 2008 using the IRS inflation parameters, so that the EITC kinks are perfectly aligned for all years.

Two elements are worth noting in Figure 3. First, there is a clear clustering of tax filers around the first kink point of the EITC. In both panels, the density is maximum exactly at the first kink point. The fact that the location of the first kink point differs between EITC recipients with one child, versus those with two or more children, constitutes strong evidence that the clustering is driven by behavioral responses to the EITC as predicted by the standard model. Second, however, we cannot discern any
systematic clustering around the second kink point of the EITC. Similarly, we cannot discern any gap in the distribution of earnings around the concave kink point where the EITC is completely phased-out. This differential response to the first kink point, versus the other kink points, is surprising in light of the standard model predicting that any convex (concave) kink should produce bunching (gap) in the distribution of earnings.

In Figure 4, we break down the sample of earners into those with nonzero self-employment income versus those zero self-employment income (and hence whose

**Figure 4. Earnings Density and the EITC: Wage Earners versus Self-Employed**

Notes: The figure displays the kernel density of earnings for wage earners (those with no self-employment earnings) and for the self-employed (those with nonzero self-employment earnings). Panel A reports the density for tax filers with one dependent child and panel B for tax filers with two or more dependent children. The charts include all years 1995–2004. The bandwidth is $400 in all kernel density estimations. The fraction self-employed in 16.1 percent and 20.5 percent in the population depicted on panels A and B (in the data sample, the unweighted fraction self-employed is 32 percent and 40 percent). We display in dotted vertical lines around the first kink point the three bands used for the elasticity estimation with $\delta = 1,500$. 

Source: Saez (2010), p. 192
systematic clustering around the second kink point of the EITC. Similarly, we cannot discern any gap in the distribution of earnings around the concave kink point where the EITC is completely phased-out. This differential response to the first kink point, versus the other kink points, is surprising in light of the standard model predicting that any convex (concave) kink should produce bunching (gap) in the distribution of earnings.

In Figure 4, we break down the sample of earners into those with nonzero self-employment income versus those zero self-employment income (and hence whose
Earnings Distributions in Lowest and Highest Bunching Deciles

Source: Chetty, Friedman, and Saez NBER'12
Fraction of Tax Filers Who Report SE Income that Maximizes EITC Refund in 1996

Source: Chetty, Friedman, and Saez NBER'12
Fraction of Tax Filers Who Report SE Income that Maximizes EITC Refund in 1999

[Map showing the fraction of tax filers with maximized EITC refund by county.]

Source: Chetty, Friedman, and Saez NBER'12
Fraction of Tax Filers Who Report SE Income that Maximizes EITC Refund in 2002

Source: Chetty, Friedman, and Saez NBER'12
Fraction of Tax Filers Who Report SE Income that Maximizes EITC Refund in 2005

Source: Chetty, Friedman, and Saez NBER'12
Fraction of Tax Filers Who Report SE Income that Maximizes EITC Refund in 2008

Source: Chetty, Friedman, and Saez NBER'12
Is the EITC having an effect on this distribution?

Source: Chetty, Friedman, and Saez NBER'12
Income Distribution For Single Wage Earners with One Child
High vs. Low Bunching Areas

Source: Chetty, Friedman, and Saez NBER'12
Earnings Distribution in the Year Before First Child Birth for Wage Earners

Source: Chetty, Friedman, and Saez NBER'12
Earnings Distribution in the Year of First Child Birth for Wage Earners

Source: Chetty, Friedman, and Saez NBER'12
Fig. 2. Total consumption: single mothers, 1984–2000.


Source: Meyer and Sullivan 2004
Relative Consumption: single women with/without children

Fig. 3. Relative total consumption: single mothers vs. single women without children, 1984–2000.

Source: Meyer and Sullivan 2004
judge allowance rate in the other cases a judge has handled. We note the judge leniency measure is calculated from all cases the judge has ever handled, not just the cases in our estimation sample. On average, each judge has handled a total of 380 cases. The mean of the leniency variable is .15 with a standard deviation of .06. The histogram reveals a wide spread in judge leniency, with approximately 22% of cases allowed by a judge at the 90th percentile compared to approximately 9% at the 10th percentile.

**Figure 3: Effect of Judge Leniency on Parents (First Stage) and Children (Reduced Form).**

Notes: Baseline sample, consisting of parents who appeal an initially denied DI claim during the period 1989-2005 (see Section 3 for further details). There are 14,893 individual observations and 79 different judges. Panel (A): Solid line is a local linear regression of parental DI allowance on judge leniency. Panel (B): Solid line is a local linear regression of child DI receipt on their parent’s judge leniency measure. All regressions include fully interacted year and department dummies. The histogram of judge leniency is shown in the background of both figures (top and bottom 0.5% excluded from the graph).

Source: Dahl, Kostol, Mogstad (2013)

Panel A shows the effect of judge leniency on a parent’s allowance rate. The graph is a flexible analog to the first stage equation (4), where we plot a local linear regression of actual parental allowance against judge leniency. The parental allowance rate is monotonically increasing in our leniency measure, and is close to linear. A one percentage point increase in the judge’s allowance rate in other cases is associated with an almost one percentage point increase in the probability the parent’s case is allowed. Panel B plots the reduced form effect of a parent’s judge leniency measure against their child’s DI participation, again using a local linear regression. The child’s DI rate is monotonically increasing in the leniency measure as well. Approximately two and a half percent of children whose parents had a relatively strict judge (leniency measure = .09, the 10th percentile) are predicted to participate in DI five years later. This can be contrasted with roughly three percent of children whose parents had a relatively lenient judge (leniency measure = .22, the 90th percentile).
EITC Schedule in 2017

0 children

1 child

2 children

3+ children

0

2000

4000

6000

Annual Credit (USD)

0

10000

20000

30000

40000

50000

60000

Earnings (USD)

0

10000

20000

30000

40000

50000

60000
EITC Maximum Credit Over Time

Source: Kleven (2018)
Labor Force Participation of Single Women
With and Without Children

Source: Kleven (2018)
Labor Force Participation of Single Women
With and Without Children

50 years of relative stability, apart from these 5 years

Source: Kleven (2018)
Labor Force Participation of Single Women
With and Without Children

14.5pp
50 years of relative stability, apart from these 5 years
2.6 4.6

Unemployment Rate
50 60 70 80 90 ...

Labor Force Participation (%)
68 70 72 74 76 78 80 82 84 86 88 90 92 94 96 98 00 02 04 06 08 10 12 14 16 18

Year
With Children Without Children

14.5pp
14pp

Source: Kleven (2018)
Labor Force Participation of Single Women
With and Without Children

Tax Reduction Act of 1975
TRA86 OBRA90 OBRA93 ARRA

Unemployment Rate
50 60 70 80 90 100

Labor Force Participation (%)
68 70 72 74 76 78 80 82 84 86 88 90 92 94 96 98 00 02 04 06 08 10 12 14 16 18

Year

With Children Without Children

Source: Kleven (2018)
Labor Force Participation of Single Women

With and Without Children

Source: Kleven (2018)
Labor Force Participation of Single Women
With and Without Children

Source: Kleven (2018)
Labor Force Participation of Single Women
With and Without Children

Source: Kleven (2018)
Labor Force Participation of Single Women
By Number of Children

Source: Kleven (2018)
Labor Force Participation of Single Women
By Number of Children

Much larger increase by those with 3+ kids

Source: Kleven (2018)
Labor Force Participation of Single Women
By Number of Children

Tax Reduction Act of 1975 TRA86 OBRA90 OBRA93 ARRA
But no increase here
by those with 3+ kids

40 50 60 70 80 ...

Year

0 children 1 child 2 children 3+ children

Source: Kleven (2018)
The first term on the right-hand side includes event-time dummies, the second term includes age dummies (to control for life cycle trends), and the third term includes year dummies (to control for time trends). We omit the event-time dummy at \( t = -1 \), implying that the event-time coefficients measure the impact of children relative to the year just before the first childbirth. We are able to identify the effects of all three sets of dummies because, conditional on age and year, there is variation in event time driven by variation in the age at which individuals have their first child. Kleven, Landais, and Søgaard (forthcoming) lay out the identification assumptions underlying this approach, compare its results to alternative approaches in the literature, and provides evidence of its ability to identify the causal effect of parenthood.

Our main outcome variable is gross labor earnings, excluding taxes or transfers, specified in levels.\(^3\) We convert the estimated level effects into percentages by calculating \( P_t^g \equiv \hat{\alpha}_t^g / E \left[ \hat{Y}_{ist}^g | t \right] \) where \( \hat{Y}_{ist}^g \) is the predicted outcome when omitting the contribution of the event dummies.\(^4\) Having estimated the impacts of children on women and men separately, we define the child penalty at event time \( t \) as \( P_t \equiv (\hat{\alpha}_t^m - \hat{\alpha}_t^w) / E \left[ \hat{Y}_{ist}^g | t \right] \). This measures the percentage by which women are falling behind men due to children.

II. Child Penalties: Results

Figures 1–3 show the effects of parenthood on earnings across the different countries. The results confirm that the existence of large child penalties is a pervasive phenomenon. In each country, the earnings of men and women evolve similarly before parenthood—after adjusting for life cycle and time trends—but diverge sharply after parenthood. Women experience a large, immediate and persistent drop in earnings after the birth of their first child, while men are essentially unaffected. Ten years after childbirth, women have not recovered and at this point the series have plateaued.

Despite these similarities, the graphs also reveal some striking differences. First, the size of the long-run child penalty (defined as the average penalty from event time five to

---

\(^3\)We specify equation (1) in levels rather than in logs to be able to keep the zeros in the data (due to nonparticipation). In the online Appendix, we present separate results on the extensive margin impacts of children.

\(^4\)To be precise, we define \( \hat{Y}_{ist}^g \equiv \sum_k \hat{\gamma}_k^g \cdot 1[k = age_{it}] + \sum_p \hat{\gamma}_p^g \cdot 1[p = i] \). Hence, \( P_t^g \) captures the year-\( t \) effect of children as a percentage of the counterfactual outcome absent children.
The first term on the right-hand side includes event-time dummies, the second term includes age dummies (to control for life cycle trends), and the third term includes year dummies (to control for time trends). We omit the event-time dummy at $t = -1$, implying that the event-time coefficients measure the impact of children relative to the year just before the first childbirth. We are able to identify the effects of all three sets of dummies because, conditional on age and year, there is variation in event time driven by variation in the age at which individuals have their first child. Kleven, Landais, and Søgaard (forthcoming) lay out the identification assumptions underlying this approach, compare its results to alternative approaches in the literature, and provides evidence of its ability to identify the causal effect of parenthood.

Our main outcome variable is gross labor earnings, excluding taxes or transfers, specified in levels.\(^3\) We convert the estimated level effects into percentages by calculating $P_t^g \equiv \hat{\alpha}_t^g / E[Y_{ist}^g | t]$ where $Y_{ist}^g$ is the predicted outcome when omitting the contribution of the event dummies.\(^4\) Having estimated the impacts of children on women and men separately, we define the child penalty at event time $t$ as $P_t \equiv (\hat{\alpha}_t^m - \hat{\alpha}_t^w) / E[Y_{ist}^g | t]$. This measures the percentage by which women are falling behind men due to children.

II. Child Penalties: Results

Figures 1–3 show the effects of parenthood on earnings across the different countries. The results confirm that the existence of large child penalties is a pervasive phenomenon. In each country, the earnings of men and women evolve similarly before parenthood—after adjusting for life cycle and time trends—but diverge sharply after parenthood. Women experience a large, immediate and persistent drop in earnings after the birth of their first child, while men are

\(^3\)We specify equation (1) in levels rather than in logs to be able to keep the zeros in the data (due to nonparticipation). In the online Appendix, we present separate results on the extensive margin impacts of children.

\(^4\)To be precise, we define $Y_{ist}^g \equiv \sum_k \beta_k^g \cdot 1[y = k] + \sum_k \gamma_k^t \cdot 1[y = k]$. Hence, $P_t^g$ captures the year-$t$ effect of children as a percentage of the counterfactual outcome absent children.

Note: See the notes to Figure 1.

FIGURE 1. CHILD PENALTIES IN EARNINGS IN SCANDINAVIAN COUNTRIES

Notes: The figure shows percentage effects of parenthood on earnings across event time $t$ for each gender $g$, i.e., $P_t^g$ defined above. The figure also displays long-run child penalties, defined as the average penalty $P_t$ from event time five to ten. Earnings are unconditional on employment status and the effects therefore include both the extensive and intensive margins.

FIGURE 2. CHILD PENALTIES IN EARNINGS IN ENGLISH-SPEAKING COUNTRIES

Source: Kleven et al. AEA-PP 2019
we see that these countries feature less dramatic short-run effects, but that the effects are growing over time.

In general, the earnings penalties can come from three margins: the extensive margin of labor supply (employment), the intensive margin of labor supply (hours worked), and the wage rate. In the online Appendix, we provide evidence on child penalties along the extensive margin. While parenthood reduces female employment everywhere, the importance of this margin varies across countries. In the Scandinavian and Germanic countries, the extensive margin effects are significantly smaller than the earnings effects, implying that a substantial fraction of the earnings penalty is driven by the intensive margin and wage-rate effects. In the United States and the United Kingdom, the employment penalty is much closer in magnitude to the earnings penalty, suggesting that the extensive margin is a key driver of penalties in those countries.

III. Child Penalties: Explanations

One set of explanations for the differences in child penalties focus on government policies. These include taxes, transfers, and family policies such as parental leave and childcare provision that directly affect mothers’ incentive to work. There is a voluminous literature on the impact of such policies on female labor supply and gender gaps (see Olivetti and Petrongolo 2017 for a review). Of particular relevance, Kleven et al. (2019) considers the impacts of parental leave and public childcare on the dynamics of child penalties. Their setting is Austria, a country where the combination of rich administrative data and a series of parental leave reforms and childcare expansions allow for compelling quasi-experimental analyses of these questions.

---

5 Angelov, Johansson, and Lindahl (2016) estimate child penalties for Sweden using a different event-study specification. An advantage of implementing the same specification across countries is that it allows for direct comparisons. The fact that Denmark and Sweden are so different is a priori surprising. We note that our earnings measure in general includes any (non-mandated) parental leave benefits paid by the employer, implying that cross-country comparisons partly reflect variation in such benefits. While employer-provided parental leave benefits do tend to be higher in Denmark than in Sweden, this is likely to have a modest impact on the relative child penalties for two reasons. One is that such employer-provided benefits were relatively small during the period we study (in Denmark we are considering first child births between 1985–2003), and the other is that those benefits are provided only during event times 0 and 1.

6 Since we do not condition our samples on having only one child, the long-run child penalties will include the effects of subsequent children and therefore depend on total fertility. However, differential fertility is unlikely to drive the variation in child penalties across countries. For example, the German-speaking countries exhibit the largest penalties despite being characterized by the lowest realized fertility at event time ten. See Table A.I in the online Appendix for descriptive statistics in each country.
Figure 4: Secondary Job Holding Rates by Secondary Earnings Level

Source: Tazhildinova (2019)

(a) same axis

(b) different axis

Notes: This figure shows the share of individuals with secondary jobs paying less than €400 per month, paying between €400 and €1000, or more than €1000 per month. The vertical red line identifies the 2003 tax reform. Source: Sample of Integrated Labour Market Biographies (SIAB) 1975 - 2010, Nuremberg 2013.
Employment Rates of Men by Age, 2019

Source: Saez AEA-PP'21

Source: OECD database online. Employment to population ratios.
Employment Rates of Women by Age, 2019

Source: OECD database online. Employment to population ratios.

Source: Saez AEA-PP’21
Employment Rates of Men and Women, aged 25-54

Source: Saez AEA-PP'21

Source: OECD database online.
Employment Rates of Men and Women, aged 25-54

Source: Saez AEA-PP’21

Source: OECD database online.
US female labor force participation, age 16-64

Source: Saez AEA-PP’21

25% increase in 1943-1945 during WW2 planned economy

Average Annual Hours of Work of Employees

Source: Saez AEA-PP’21

US has 40 hour/week and no mandatory paid vacation

1968: 4th week of paid vacation

1982: 5th week + 39 hours/week

2000-2: 35 hours/week

Source: OECD database online. Includes all ages, genders, and part-time, full-time, overtime.
Negative Income Tax Experiment

\[ c = z - T(z) \]

- **NIT Treatment:** Transfer \( G \) phased-out with earnings \( z \) at tax rate \( \tau \)

- **Control group:** slope = 1

- **Slope:** \( 1 - \tau \)

- **Point of Interest:** \( z^* \)

- **45° Line**

- **Graph Axes:**
  - Y-axis: \( c \)
  - X-axis: pre-tax income \( z \)
Negative Income Tax Experiment

\[ c = z - T(z) \]

NIT Treatment
Negative income and substitution effects on \( z \)

slope = \( 1 - \tau \)

Control group: slope = 1

Pre-tax income \( z \)
EITC and intensive labor supply

\[ c = z - T(z) \]

Budget with EITC
EITC and intensive labor supply

c = z - T(z)

Negative income and substitution effects on z

Negative income effects on z

Positive substitution effects on z

pre-tax income z

0

45°
SNAP Tracks Changes in Share of Population Near or Below the Poverty Line

Share of individuals with income at or below 130% of poverty line

Share of individuals participating in SNAP

Note: Poverty estimates are annual estimates. SNAP shares of resident population are calendar year averages.

Sources: U.S. Census Bureau, U.S. Department of Agriculture
Figure 1: Long-Run Evolution of EITC and Cash Welfare

Source: Internal Revenue Service (EITC) and Department of Health and Human Services (AFDC/TANF).

Notes: The red series show the annual number of federal EITC recipients between 1966-2016. The blue series show the average monthly number of Aid to Families with Dependent Children (AFDC) recipients between 1966-1996, and the average monthly number of Temporary Assistance for Needy Families (TANF) recipients between 1997-2016.
Means-tested Transfers in the US, 1960-2019

but the Act also stated that no county could run both the FSP and the CDP. Thus, for counties which previously ran a CDP, adoption of the FSP implies termination of the CDP.\footnote{This transition in nutritional assistance would tend to bias downward FSP impact estimates, but we do not think this bias is substantial because of the limited scope of the CDP. The CDP was not available in all counties and recipients often had to travel long distances to pick up the items. Further, the commodities were distributed infrequently and inconsistently, and provided a very narrow set of commodities—the most frequently available were flour, cornmeal, rice, dried milk, peanut butter, and rolled wheat (Citizens’ Board of Inquiry 1968). In contrast, food stamp benefits can be used to purchase a wide range of grocery food items.} The political accounts of the time suggest that debates about adopting the FSP pitted powerful agricultural interests (who favored the CDP) against advocates for the poor (who favored the FSP: see MacDonald 1977; Berry 1984). In particular, counties with strong support for farming interests (e.g., Southern or rural counties) may be late adopters of the FSP. On the other hand, counties with strong support for the low-income population (e.g., Northern, urban counties with large poor populations) may adopt FSP earlier in the period. This systematic variation in food stamp adoption could lead to spurious estimates of the program impact if those same county characteristics are associated with differential trends in the outcome variables.

In earlier work (Hoynes and Schanzenbach 2009), we documented that larger counties with a greater fraction of the population that was urban, black, or low income indeed implemented the FSP earlier (i.e., consistent with the historical accounts).\footnote{For more detail, see Table 1 in Hoynes and Schanzenbach (2009).} Nevertheless, we found that the county characteristics explain very little of the variation in adoption dates. This is consistent with the characterization of funding limits controlling the movement of counties off the waiting list to start up

\begin{figure}
\centering
\includegraphics[width=\textwidth]{food_stamp_program_start_date.png}
\caption{Food Stamp Program Start Date, by County, 1961–1974}
\textit{Notes:} Authors’ tabulations of food stamp administrative data (US Department of Agriculture, various years). The shading corresponds to the county FSP start date, where darker shading indicates later county implementation.
\end{figure}
Amendments to the Food Stamp Act, which mandated that all counties offer FSP by 1975. Figure 1 plots the percent of counties with a FSP from 1960 to 1975. During the pilot phase (1961–1964), FSP coverage increased slowly. Beginning in 1964, program growth accelerated; coverage expanded at a steady pace until all counties were covered in 1974. Furthermore, there was substantial heterogeneity in timing of adoption of the FSP, both within and across states. The map in Figure 2 shades counties according to date of FSP adoption (darker shading denotes a later start-up date). Our basic identification strategy considers the month of FSP adoption for each county to trigger the beginning of the FSP “treatment.”

For our identification strategy to yield causal estimates of the program, it is important to establish that the timing of FSP adoption appears to be exogenous and here we summarize what we have examined in our earlier work (Hoynes and Schanzenbach 2009). Prior to the FSP, some counties provided food aid through the commodity distribution program (CDP) which took surplus food purchased by the Federal government as part of an agricultural price support policy and distributed those goods to the poor. The 1964 Food Stamp Act allowed for counties to voluntarily set up a FSP.

Counts are weighted by their 1970 population. Note this is not the food stamp caseload, but represents the percent of the US population that lived in a county with a FSP. Online Appendix Figure 1 reproduces this figure and adds the county-level coverage rate using the PSID data. The data available in the PSID line up well with the national rollout trends.
given that we find zero impact on economic self-sufficiency for men (Table 6). Note
that these are the reverse of a typical event study graph, in that negative “event time”
is the case where a person was fully treated (food stamps was in place in their county
prior to birth). Further, treatment (exposure to the program) increases as we move
from the right (treated in later life) to the left (treated in early life). Finally, as we
have said before, once the treatment turns on it does not turn off.

While we do not have a strong prediction about the precise shape of the treatment
effects, our hypothesis is that the impact of the FSP treatment should decline as age
at initial exposure increases. Or to state the reverse, the younger the initial age of
exposure the larger the (cumulative) effect of the FSP. If exposure in later childhood
does not matter, then the event study coefficients should be flat on the right end of
the graph (suggesting no “pretrend”). Eventually, once we hit the point in early
childhood when exposure matters, a movement left (toward earlier initial exposure)
should reduce the metabolic syndrome index (or increase economic self-sufficiency).
Eventually, the event study should be flat once exposure is “complete” (exposure is
prior to conception or an event time of −1 or before).

The results in Figure 3 are highly consistent with these predictions and quite
encouraging for our research design. They show that the largest effects of the food
stamp treatment (in this case a reduction in metabolic syndrome is good and so a

Figure 3. Event Study Estimates of the Impact of FSP Exposure on Metabolic Syndrome Index
(High Participation Sample)

Notes: The figure plots coefficients from an event-study analysis. Event time is defined as age when FSP is imple-
mented in the birth county. The models are estimated for the sample of individuals born into families where the head
has less than a high school education. Age 10–11 is the omitted year so estimates are relative to that point. See the
text for a description of the model.
medical criteria to determine eligibility at age 18, whereas previously age 18 reviews only required that children not demonstrate improvement in their medical condition to pass the review. The definition of disability is different for adults and children: for children, disability is defined in terms of age-appropriate activity, whereas for adults disability is defined as an inability to work. Conditions like ADHD may qualify a child for SSI but are less likely to qualify an adult unless it is judged to be severe enough to preclude work. Thus the new PRWORA requirement effectively made the age 18 review criteria more stringent.

Deshpande (2016) documents that the PRWORA rules were applied to children with an 18th birthday after August 22, 1996, the date of enactment. Figure II replicates, for the CJARS-coverage sample used in this article, the first-stage result from Deshpande (2016) showing that almost no children with an 18th birthday before PRWORA enactment received an age 18 medical review, while nearly 90% of children with an 18th birthday after PRWORA enactment received a review. (There is a decline in
DOES WELFARE PREVENT CRIME?

FIGURE III

Reduced Form: Criminal Justice Outcomes across Cutoff

The figure plots the total number of charges between ages 18 and 38 (top left); the average annual likelihood of incarceration from ages 18 to 38 (top right), defined as the likelihood of incarceration in each year from age 18 to 38 averaged across the 21 years (where the indicator is turned on in any year with at least one day spent in prison); the total number of income-generating charges (bottom left); and the total number of non-income-generating charges (bottom right). Outcomes are residualized (and then control mean added back) using the same covariates used in the main specification: medical diary (severity), sex, first age of SSI receipt, body system code, family structure, parent earnings, and state. The sample is SSI children with an 18th birthday within 18 months of the August 22, 1996, cutoff who reside in a county with CJARS coverage. Table I reports point estimates and standard errors.

Importantly, these results mean that the two main CJARS data sources—court charges and corrections—independently show a large and precise increase in criminal justice involvement across the cutoff.11

11. What explains the upward trend in number of charges by birthdate? The answer appears to be a secular increase in criminal charges for younger birth cohorts. Online Appendix Figure B5 shows an upward trend in criminal charges and incarceration for the general population with the same birthdate range.