ESSAYS ON INCOME, SOCIAL POLICY, AND EDUCATION

By

Michelle Maxfield

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Economics—Doctor of Philosophy

2014

ABSTRACT

ESSAYS ON INCOME, SOCIAL POLICY, AND EDUCATION

By

Michelle Maxfield

Chapter 1: "The Effects of the Earned Income Tax Credit on Child Achievement and Long-Term Educational Attainment." The Earned Income Tax Credit (EITC) is a significant source of government assistance to low income families. Total outlay reached over $50 billion in 2008, with more than 97 percent of aid received by families with children (Internal Revenue Service 2011). Despite its size and pro-child goals, relatively little is known about how the EITC affects children directly. This study directly links EITC receipt throughout all ages of childhood to both contemporaneous achievement and long-run educational attainment. I take advantage of both Federal tax code changes and state EITC adoptions, which result in large variation in EITC generosity across state, time, and family size. Using the 1979 National Longitudinal Survey of Youth, I find that EITC expansions improve both contemporaneous and long-run educational outcomes of children. An increase in the maximum EITC of $1,000 (2008 dollars) in a given year significantly increases math achievement by about 0.072 nationally normed standard deviations. This change in EITC generosity during childhood also increases the probability of graduating high school or receiving a GED at age 19 by about 2.1 percentage points and increases the probability of completing one or more years of college by age 19 by about 1.4 percentage points. Estimated effects are larger for boys and minority children, and I find evidence that an expansion in the EITC is more effective at improving educational outcomes for children who are younger during the expansion.

Chapter 2: “The Effects of the Earned Income Tax Credit on Net Family Financial Resources.” This study is the first of my knowledge to examine the effects of EITC expansions on a comprehensive measure of total net family income as well as the various income sources that comprise this measure. Using the Current Population Survey, I find an increase in labor force participation of about 1.7 percentage points and a 1.3 percentage point decline in the poverty rate following a relative increase in the maximum EITC of about $1,900 (2008 dollars) for low income families. This EITC expansion also increases relative earnings by about $471, increases EITC payments by about $742, and increases total net family income by about $527. I also find larger impacts for single mothers and minority families, suggesting that the program is well-targeted at the most disadvantaged families.

Chapter 3: “An Evaluation of Empirical Bayes’ Estimation of Value-Added Teacher Performance Measures.” Empirical Bayes’ (EB) estimation is a widely used procedure to calculate teacher value-added. It is primarily viewed as a way to make imprecise estimates more reliable. In this paper, we review the theory of EB estimation and use simulated data to study its ability to properly rank teachers. We compare the performance of EB estimators with that of other widely used value-added estimators under different teacher assignment scenarios. We find that, although EB estimators generally perform well under random assignment of teachers to classrooms, their performance generally suffers under non-random teacher assignment. Under non-random assignment, estimators that explicitly (if imperfectly) control for the teacher assignment mechanism perform the best out of all the estimators we examine. We also find that shrinking the estimates, as in EB estimation, does not itself substantially boost performance.

ACKNOWLEDGEMENTS

The author is grateful to Stacy Dickert-Conlin, Leslie Papke, Jeffrey Wooldridge, Steven Haider, and Sara Witmer for their guidance and comments. The research presented here is partially supported by the Institute of Education Sciences, U.S. Department of Education, through grant R305B090011 to Michigan State University. The opinions expressed are those of the author and do not represent the views of the Institute or the U.S. Department of Education.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

The Effects of the Earned Income Tax Credit on Child Achievement and Long-Term Educational Attainment

1.1 Introduction

The Earned Income Tax Credit (EITC) is a significant source of government aid to needy families and has grown dramatically since its inception in 1975. Total outlay reached over $50 billion in 2008, with more than 97 percent of aid received by families with children (Internal Revenue Service 2011). The largest expansion in the EITC came as part of President Bill Clinton's Omnibus Budget Reconciliation Act (OBRA) in 1993. At that time, more than 12 million children, one in every four, were living in poverty, making up about one third of the total population living in poverty in the United States (Mink 1993). The EITC gained support from many child advocates, including the National Commission on Children (1993), as the president's plan ensured that no family with a parent working full-time would live below the poverty line (Stupak 1993). In a congressional session addressing children's initiatives, Congresswoman Karen Shepard (1993) stated: "If you believe that work should be rewarded and that children deserve security, you should support expanding the Earned Income Tax Credit. Plain and simple."

Despite the size and pro-child goals of the EITC, relatively little is currently known about how the credit affects children directly. Until recently, studies focused only on indirect measures of child well-being such as poverty, parental labor supply, marriage, fertility, and consumption (see Hotz and Scholz 2003 and Eissa and Hoynes 2006 for reviews of the literature). Without knowing the direct impacts of the EITC on child outcomes such as physical and mental health, cognition, and long-run economic sufficiency, it is difficult to accurately assess the performance of the program. This paper looks to address the effects of the EITC on both contemporaneous and long-term educational outcomes.

Only three studies I am aware of directly examine the effects of the EITC on child cognitive outcomes. Dahl and Lochner (2012) use the EITC as an exogenous source of income variation to determine the effects of family income on child achievement in math and reading. Chetty et al. (2011a) first analyze the effects of the EITC and the Child Tax Credit on math and reading scores. They then use the finding that the credits improve test scores to consider possible long-term effects by examining how test score gains from being assigned a more effective teacher affect outcomes such as college attendance and earnings. Michelmore (2013) examines state EITCs as an unexplored source of financial aid to determine effects of income on educational attainment.

Using the National Longitudinal Survey of Youth (NLSY), which follows mothers and their children over time, I am able to directly estimate the impact of exposure to a more generous EITC during childhood on both contemporaneous achievement and long-run educational attainment for children of all ages. These data also allow me to estimate changes to family resources following EITC expansion, giving some insight into the mechanisms behind the effects on child outcomes. I take advantage of both Federal tax code changes and state EITC adoptions, which result in large variation in EITC generosity across state, time, and family size.

I find that the EITC is an effective policy for improving both contemporaneous and long-run educational outcomes of children. I estimate that OBRA 1993, which increased the Federal maximum EITC payment by about $3,000 (2008 dollars), had large, significant effects on children. For an elementary-aged child in a family with 2 or more children, OBRA 1993 increased math achievement by about 0.215 nationally normed standard deviations, increased probability of graduating high school or receiving a GED at age 19 by about 7.2 percentage points, and increased probability of completing one or more years of college by age 19 by about

4.8 percentage points. Along with changes to educational outcomes, OBRA 1993 resulted in other changes in the household, including an increase in net family income inclusive of EITC and welfare payments of about $2,664 and an increase in maternal labor force participation.

In the following section, I review the institutional details of the EITC. Section 1.3 outlines how the EITC might affect children and reviews the previous literature on this topic. Section 1.4 describes the NLSY data and presents summary statistics for my sample. Section 1.5 details my empirical strategy, and Section 1.6 presents the results. I summarize the findings and conclude in Section 1.7.

1.2 Institutional Details of the EITC

The EITC began in 1975 with modest credits for low income families with children as a way to offset payroll taxes. Since then, the Federal government expanded the EITC multiple times in an effort to create an anti-welfare, anti-poverty, and pro-work tool (Ventry 2000). The credit is refundable and only available to families who work. It is based on a family's earned income, number of children, and state of residence. Table 1 shows the Federal EITC parameters for the years I examine, 1987 to 2000. As the table illustrates, there is an initial "phase-in" range and rate, where the credit is equal to the subsidy rate times the family's earned income until the maximum credit is reached. The family then receives the maximum credit during the "flat" range. Once a family reaches a certain level of income, they enter a "phase-out" range, where the credit is reduced at the phase-out rate. Thus, only families below a certain level of income are eligible for the credit in each year. Families are given the option to receive the credit with periodic payments throughout the year as opposed to a one-time lump sum. However, less than five percent of families exercised this option during the time frame I study (Friedman 2000).

TABLE 1
Federal EITC Parameters, 1987-2000

Calendar year	Credit rate (%)	Min income for max credit	Max credit	Phase-out rate (%)	Phase-out range Beginning income	Phase-out range Ending income
1987	14	6,080	851	10	6,920	15,432
1988	14	6,240	874	10	9,840	18,576
1989	14	6,500	910	10	10,240	19,340
1990	14	6,810	953	10	10,730	20,264
1991						
One child	16.7	7,140	1,192	11.93	11,250	21,250
Two children	17.3	7,140	1,235	12.36	11,250	21,250
1992						
One child	17.6	7,520	1,324	12.57	11,840	22,370
Two children	18.4	7,520	1,384	13.14	11,840	22,370
1993						
One child	18.5	7,750	1,434	13.21	12,200	23,050
Two children	19.5	7,750	1,511	13.93	12,200	23,050
1994						
No children	7.65	4,000	306	7.65	5,000	9,000
One child	26.3	7,750	2,038	15.98	11,000	23,755
Two children	30	8,425	2,528	17.68	11,000	25,296
1995						
No children	7.65	4,100	314	7.65	5,130	9,230
One child	34	6,160	2,094	15.98	11,290	24,396
Two children	36	8,640	3,110	20.22	11,290	26,673
1996						
No children	7.65	4,220	323	7.65	5,280	9,500
One child	34	6,330	2,152	15.98	11,610	25,078
Two children	40	8,890	3,556	21.06	11,610	28,495
1997						
No children	7.65	4,340	332	7.65	5,430	9,770
One child	34	6,500	2,210	15.98	11,930	25,750
Two children	40	9,140	3,656	21.06	11,930	29,290
1998						
No children	7.65	4,460	341	7.65	5,570	10,030
One child	34	6,680	2,271	15.98	12,260	26,473
Two children	40	9,390	3,756	21.06	12,260	30,095
1999						
No children	7.65	4,530	347	7.65	5,670	10,200
One child	34	6,800	2,312	15.98	12,460	26,928
Two children	40	9,540	3,816	21.06	12,460	30,580
2000						
No children	7.65	4,610	353	7.65	5,770	10,380
One child	34	6,920	2,353	15.98	12,690	27,413
Two children	40	9,720	3,888	21.06	12,690	31,152

Source: Joint Committee on Taxation, Ways and Means Committee (2004).
Note: Dollar amounts unadjusted for inflation

Thus, the vast majority of families receive their EITC credit as a lump sum upon filing their tax returns, with over 80 percent of families receiving the credit by the end of March (LaLumia 2013).

In addition to Federal funding of the credit, many states have their own credits that typically "piggyback" onto the Federal credits – meaning some states will increase the Federal EITC credit by a given percentage. The states vary substantially on the generosity of their add-ons, whether they offer it to families without children, and whether the credit is refundable. Table 2 contains the state EITC parameters from 1987 to 2000, the time period covered by the data in this paper. As seen in the table, the state add-ons range from 4 to 75 percent in this time frame, and, by 2000, fifteen states enacted their own EITCs.

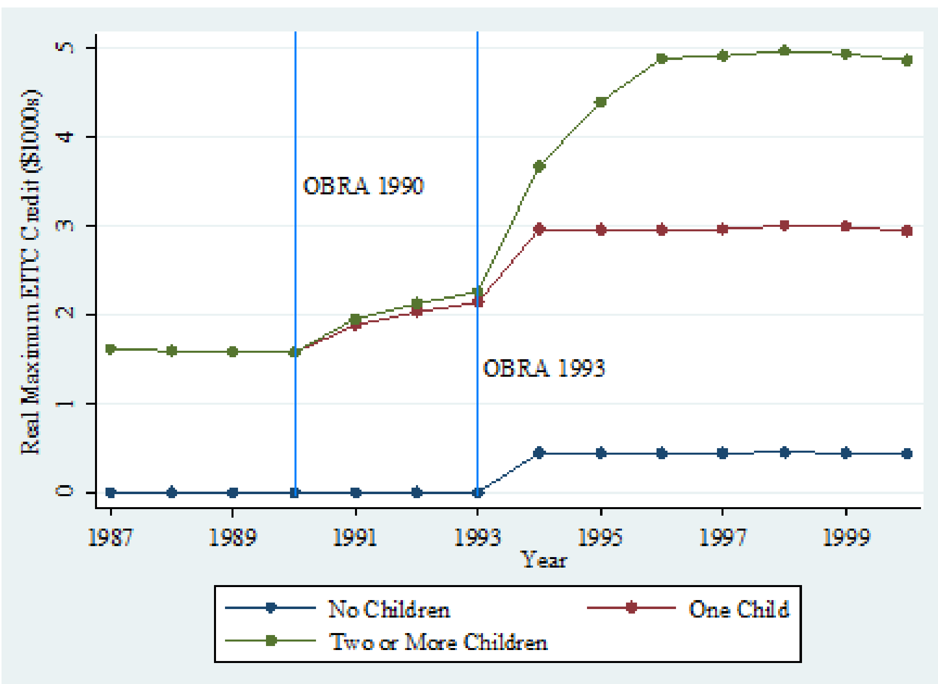
TABLE 2
State EITC Supplements, 1987-2000 (%)

State	CO	DC	IA	IL	KS	MA	MD	MD	ME	MN	MN	NJ	NY	OR	RI	VT	WI	WI	WI
No. Children	0+	0+	0+	0+	0+	0+	1+	1+	0+	0	1+	1+	0+	0+	0+	0+	1	2	3+
1987							50								23				
1988							50								23	23			
1989							50								23	25	5	25	75
1990			5				50								28	28	5	25	75
1991			6.5				50			10	10				28	28	5	25	75
1992			6.5				50			10	10				28	28	5	25	75
1993			6.5				50			15	15				28	28	5	25	75
1994			6.5				50			15	15		7.5		28	25	4.4	21	63
1995			6.5				50			15	15		10		28	25	4	16	50
1996			6.5				50			15	15		20		28	25	4	14	43
1997			6.5			10	50			15	15		20	5	28	25	4	14	43
1998			6.5		10	10	50	10		15	25		20	5	27	25	4	14	43
1999	8.5		6.5		10	10	50	10		25	25		20	5	27	25	4	14	43
2000	10	10	6.5	5	10	10	50	15	5	25	33	10	23	5	26	32	4	14	43
Refundable?	Y	Y	N	N	Y	Y	N	Y	N	Y	Y	Y	Y	N	N	Y	Y	Y	Y

Sources: Center on Budget and Policy Priorities and Leigh (2010).
Notes: No. Children is the number of children required for eligibility of the state supplement. Supplement is the percentage top-up of the federal EITC payment.

Figure 1 plots the real (2008 dollars) value of the maximum Federal EITC credit by tax year and number of children from 1987 to 2000. Two main law changes, the 1990 and 1993 enactments of the OBRA, resulted in real expansions in the Federal maximum credit. OBRA 1993 changes were quite substantial and also increased the Federal maximum EITC differentially by number of children. For example, a family with two or more children and real earnings of $12,000 in 1993 and 1996 would receive the maximum Federal EITC payment in both years of $2,251 and $4,880, respectively. Thus, the EITC increases income for this family by about 19 percent before OBRA 1993 and by about 41 percent after the law change is fully phased in.

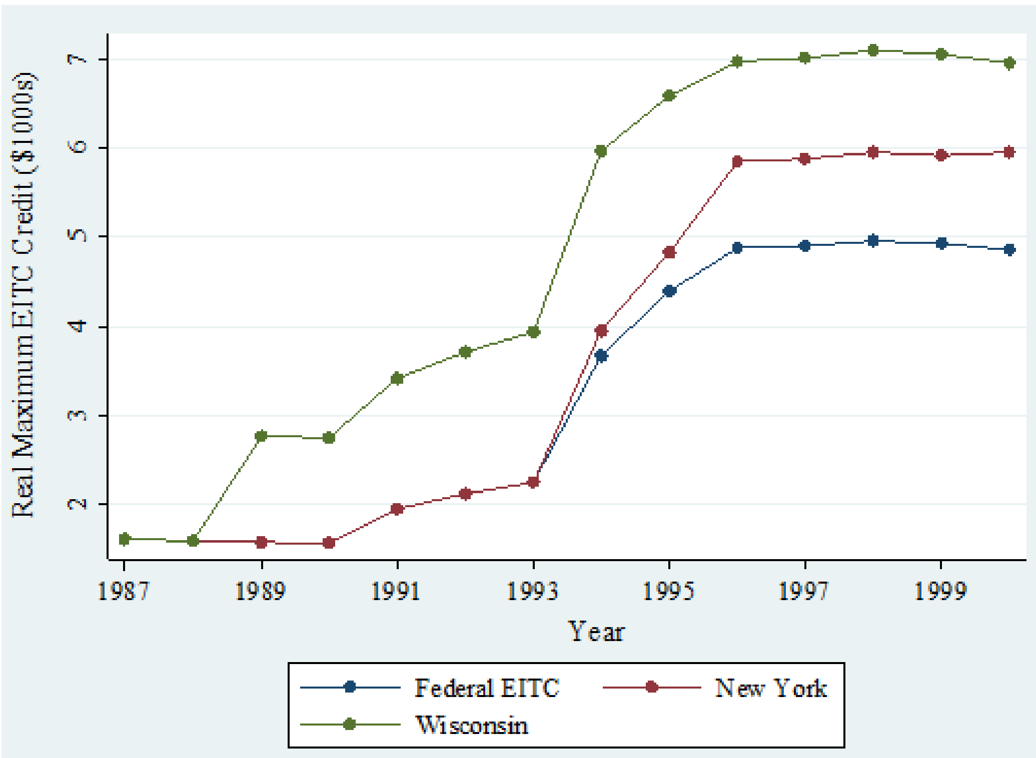
FIGURE 1
Real Maximum Federal EITC Credit by Tax Year and Number of Children (2008$), 1987-2000



Sources: Joint Committee on Taxation, Ways and Means Committee (2004).
Formatting adopted from Hoynes et al. (2012).

Figure 2 plots the real maximum Federal EITC credit for a family with three children and the combined Federal and state EITC maximum values for New York and Wisconsin from 1987 to 2000. The figure illustrates that the state EITC add-ons can be quite large as well. For example, the maximum credit in Wisconsin increased from about $1,600 to nearly $7,000 over this time period, while the Federal credit increased from about $1,600 to about $5,000.

FIGURE 2
Real Maximum EITC Credit by Tax Year (2008$), Family with 3 children



Sources: Joint Committee on Taxation, Ways and Means Committee (2004), Center on Budget and Policy Priorities, and Leigh (2010).

1.3 The EITC and Child Outcomes

1.3.1 How the EITC Affects Children

The EITC changes the home environments of children in two main ways – changes in the labor supply decisions of their mothers and changes in family income. The structure of the credit provides incentives for altering child bearing and marriage decisions as well, but previous studies have found no effect of the EITC on these outcomes (Eissa and Hoynes 2000; Ellwood 2000; Dickert-Conlin and Houser 2002; Hotz and Sholz 2003; and Baughman and Dickert-Conlin 2003 and 2009). Thus, I focus this discussion on the effects of maternal labor supply and family income.

The structure of the EITC creates different labor supply incentives depending upon the taxable income of the family. Assuming leisure is a normal good and the mother is the sole earner in the family, an EITC expansion creates an unambiguously positive incentive to enter the labor force, as it increases the potential wage of those not participating in the labor force.[1] For those mothers already participating, the incentive depends upon her income and the EITC parameters in a given year. If the mother is working and her income falls in the "phase-in" range of the EITC, there is a substitution effect away from leisure since the EITC-induced wage increase makes leisure more expensive, and there is an income effect to consume more leisure. Thus, the overall effect on hours worked is ambiguous. By similar reasoning, women in the "flat" or "phase-out" range have an unambiguous incentive to work less. Women with family income above the cutoff to be eligible for the EITC (end of the phase-out range) may also have an incentive to work less depending on their preferences and how close they are to the end of the phase-out range. As the EITC is based on family income, mothers filing jointly with a wage-earning husband will be more likely to fall in the flat or phase-out range of the EITC schedule, so these women are likely to be induced to decrease their hours worked, or possibly even leave the

[1] Technically you must also assume that the substitution effect dominates the income effect for a nonzero number of women. If the income effect dominates, the response is to stay out of the labor force.

labor force altogether (see Hotz and Scholz 2003 for a more detailed theoretical discussion of labor supply responses to the EITC).

Previous work confirms these labor supply predictions. First, EITC expansions substantially increase the labor force participation (LFP) of single mothers (Dickert et al. 1995; Eissa and Liebman 1996; Ellwood 2000; Meyer and Rosenbaum 2000 and 2001; Neumark and Wascher 2001; Grogger 2003; Hotz et al. 2006; Rothstein 2007; and Adireksombat 2010). If anything, the credit *modestly* decreases the LFP of married mothers (Dickert et al. 1995; Ellwood 2000; and Eissa and Hoynes 2004).

Evidence on the effects on hours worked for those women already in the labor force is mixed, with some studies finding no effect (Eissa and Liebman 1996; Liebman 1998; Meyer and Rosenbaum 1999; and Rothstein 2007) and others finding a slight decrease in hours worked following an EITC expansion (Dickert et al. 1995; Neumark and Wascher 2001; and Saez 2010). These mixed results likely stem from evidence that EITC recipients are not well informed of the kinked structure of the EITC (Olson and Davis 1994; Smeeding et al. 2000; Ross-Phillips 2001; Romich and Weisner 2000; Maag 2005; and Chetty and Saez 2013). Supporting this, Chetty and Saez (2013) and Chetty et al. (2013) find that there is more "bunching" of incomes at kink points in the EITC when recipients live in neighborhoods with higher levels of knowledge about the EITC.

It is not clear ex-ante how maternal labor supply itself affects children, but two main hypotheses arise in the literature. The first is that maternal LFP could be harmful, as the mother spends less time with the child. This is likely most important at very young ages of a child's life. You could also posit that less time spent with children could be beneficial, depending on the quality of the alternative care, such as other family members or daycare centers. The second hypothesis is that a working mother might provide a better example for children, changing future career expectations or aspirations, especially for girls (Goldberger at al. 2008 and Brooks-Gunn et al. 2011).

There is an expansive literature examining the relationship between maternal labor supply and child behavioral and cognitive outcomes. However, the endogeneity of maternal labor supply offers challenges to establishing causal relationships between working and child outcomes. Mothers who work have very different (more favorable) observable characteristics than those who do not work. Thus, it is likely that there is something unobservable about these mothers, like ability, intelligence, and motivation, which influence their decision to participate in the labor force (Hill et al. 2005).

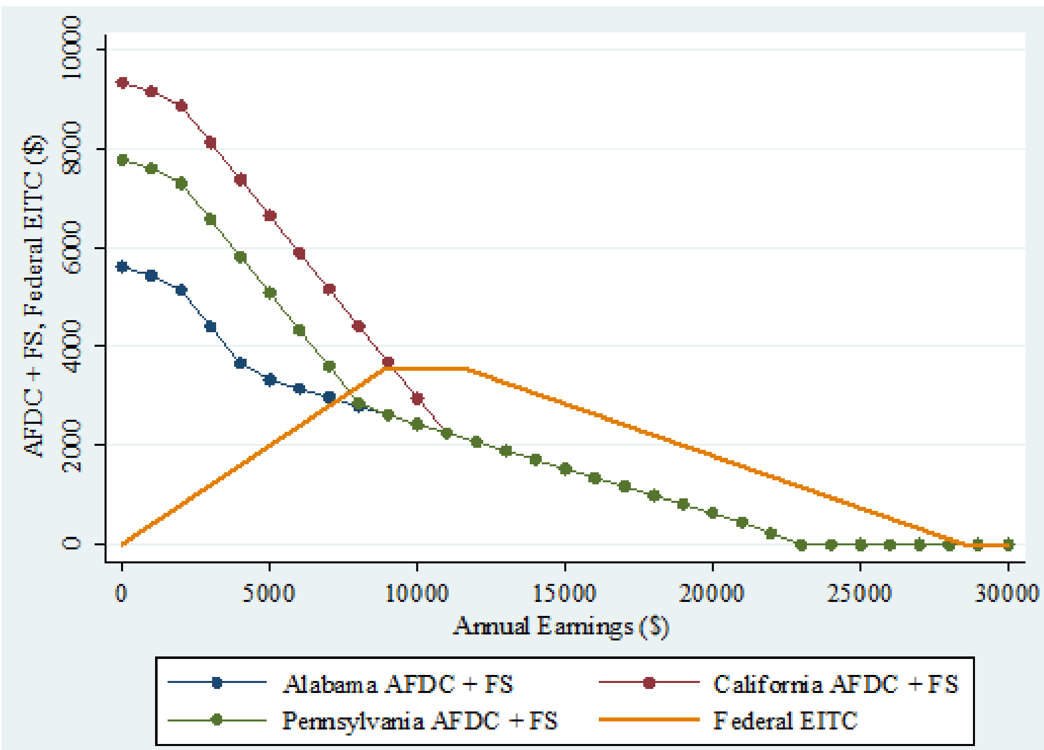
Much of the literature suggests that maternal labor supply may be harmful during early childhood, increasing behavioral issues and decreasing achievement (see Goldberger at al. 2008 and Brooks-Gunn et al. 2011 for a current review of this literature). However, the literature suggests that maternal LFP may be beneficial to child cognition beyond the first few years of a child's life. Using the NLSY, James-Burdumy (2005) uses family fixed effects and instruments for maternal labor supply using the percent of the county labor force employed in services. She finds that maternal employment in the first year of a child's life has very small negative effects on math and reading scores and that weeks worked by the mother in the third year of a child's life positively affect math scores.

As with maternal labor supply, the literature on the effects of parental income on child development is also plagued by similar endogeneity issues. As a result, most studies are correlational in nature with mixed results. Using longitudinal data from Norway, Løken et al. (2012) address the endogeneity of family income using sibling fixed effects as well as by instrumenting for income using a dummy for whether a family lived in a county that experienced an unexpected economic boom following an oil discovery. Using a quadratic specification of family income and the instrumental variables approach, they find a sizable effect of income on education attainment, high school dropout rates, and adult IQ later in life with both approaches. For a family with about $8,500 (1999 U.S. dollars) in average family income during ages one through 11, an increase in average family income of about $1,600 increases years of education by about 0.1 and decreases the probability of being a high school drop out by about 0.07.

It is possible that the income increases induced by the EITC are different from a general increase in income, as EITC payments are typically received as a lump sum once a year. Romich and Weisner (2000) provide qualitative evidence that EITC recipients in Wisconsin spend EITC funds differently than typical work income. Recipients spend the credit on housing, cars or car-related expenses, childcare, children's clothing and educational items, or paying off bills. Smeeding et al. (2000) confirms this hypothesis using data from Chicago, finding that the large majority of recipients make purchases with their EITC payment that they would not be able to if they did not receive the EITC. They find that about 80 percent of recipients expected to pay a bill or make a commodity purchase, 50 percent expected to save at least some of their refund, 16 percent planned to pay tuition, and 22 percent planned to use some funds for a car-related expenditure. The authors argue that such expenditures may improve the social mobility of these families, which could improve child development.

For a large portion of EITC recipients, earnings that lead to increased EITC receipt result in lower cash welfare and food assistance benefits, where incentives vary by state. Figure 3 depicts the 1996 annual total Aid to Families with Dependent Children (AFDC) welfare cash benefits plus annual food stamps benefits for Alabama, California, and Pennsylvania along with the Federal EITC schedule (none of these states had their own EITC this year). For example, in the phase-in region of the EITC schedule, an increase in earned income leads to an increase in EITC payments but a decrease in combined AFDC and food stamps benefits. In the absence of the EITC, the implicit tax rate on earnings is near 100 percent for the majority of AFDC and food stamps recipients (Blank 2002). The EITC helps offset this tax rate. Although income loss from welfare receipt decline would most likely harm child development, prior research suggests

FIGURE 3
1996 Benefit Schedule for AFDC, Food Stamps, and EITC
Mothers with Two Children, Alabama, California, and Pennsylvania



Sources: U.S. House of Representatives, Committee on Ways and Means (1996). Joint Committee on Taxation, Ways and Means Committee (2004). Formatting partially adopted from Meyer and Rosenbaum (2000).
Notes: I assume women are past their first four months of work, have no unearned income, and claim child care expenses equal to 20 percent of earnings up to $350 per month. I assume the standard AFDC earnings disregard of $120 per month plus the child care expenses above. I assume the standard food stamps deductions of 20 percent of earnings plus $134 per month plus the child care expenses above. Shelter expenses are ignored in food stamps calculations. AFDC payments count as income in food stamps benefit calculations but not vice versa.

that welfare receipt itself might negatively impact children due to the social stigma related to receipt (Levine and Zimmerman 2005).

The above discussion illustrates that the effect of the EITC on child development is an open empirical question, as the effects of changes to maternal employment and income caused by changes in EITC generosity could be contradictory. As mentioned above, relatively little work exists on the direct effects of the EITC on child outcomes, with existing studies focusing mainly on child health. I review the existing literature below as well as some findings from related government programs.

1.3.2 Evidence on the Effects of the EITC and Related Programs on Child Outcomes

Taking advantage of the large differential expansions in the EITC with respect to the number of children from OBRA 1993, Hoynes et al. (2012) and Baker (2008) employ difference-in-difference (DiD) techniques to estimate the effect of the EITC expansion on infant health. They both find that being exposed to a more generous EITC schedule during pregnancy reduces the likelihood of low birth weight. Strully et al. (2010) find that living in a state with an EITC supplement also increases birth weight. Baughman and Duchovny (2012) find that an increase in the maximum state EITC raises the probability that children ages 6 to 11 are in better health, but find no effects on the health outcomes of younger children.

Using the NLSY, Dahl and Lochner (2012) estimate the effects of current family income on achievement for children age 5 to 15. They instrument for changes in income using predicted changes in income based on lagged pre-tax income and changes to the Federal EITC schedule (with a flexible control function for lagged pre-tax income included as well). They find that a $1,000 (2000 dollars) increase in income leads to an increase in combined math and reading achievement of about 0.061 SD, with largest effects for reading comprehension (0.036 SD for reading recognition, 0.061 SD for reading comprehension, and 0.058 SD for math). They find larger effects for single mothers and minority children. They also find larger effects for children under age 12 compared to older children (0.077 SD and 0.052 SD, respectively) and much larger effects for boys compared to girls (0.088 SD and 0.040 SD respectively).

Chetty et al. (2011a) use the Internal Revenue Service income tax data and administrative data from a large anonymous school district to estimate the long-term effects of the EITC and Child Tax Credit (CTC). They use non-linearity in the schedule of the two tax credits to identify contemporaneous effects of tax credits on child test scores in grades three through eight (grades that are tested for accountability purposes), but their identification comes mainly from changes in the EITC. The tax data are only available beginning in 1996, so they are unable to utilize the largest changes in EITC generosity to date resulting from OBRA 1993. Also, data constraints do not allow them to directly link changes in the EITC to long-term outcomes. They proceed in two steps – first estimating the effect of tax credits on contemporaneous child test scores and then estimating the effect of test score gains on long-run outcomes using teacher assignment as exogenous variation in test scores.

They find that a $1,000 (2010 dollars) increase in tax credits in a single year raises combined math and reading achievement by about 0.080 SD, with greater effects for math than reading (0.093 SD compared to 0.062 SD). Estimated effects are larger in middle school (0.085 SD) than in elementary (0.073 SD). Since they cannot estimate the long-term effects of tax credits directly, they use the finding that tax credits improve test scores to consider possible long-term effects by examining how test score gains from being assigned a more effective teacher affect outcomes. They find that a one SD increase in test scores in a single grade raises the probability of college attendance at age 20 by about 5 percentage points (sample mean of 37 percent), improves the quality of college attended, and raises earnings at age 28 by about 9 percent. They also find that higher test scores are associated with reductions in the probability of teen pregnancy and an increase in 401(k) savings. However, as the authors point out, to make any causal inferences on the effects of tax credits on long-run outcomes you must assume that the effects of higher scores resulting from being assigned a better teacher are the same as those that would result from receiving a higher tax credit. There are many reasons these could differ including teacher cheating or teaching students only material that will be tested (i.e. "teaching to the test").

In concurrent work, Michelmore (2013) looks at the effects of state EITCs on educational attainment of children whose parents have a high school education or less (likely eligible for the EITC). She uses a DiD approach comparing 18 to 23 year olds in states with an EITC to those without an EITC, and a triple-difference specification using children from more educated households as the control group to account for state-level trends in educational outcomes. Using the Survey of Income and Program Participation, Michelmore finds that a $1,000 (2011 dollars) increase in the combined state and Federal EITC maximum increases years of schooling by 0.11, increases the probability of completing high school by 2.0 percentage points, and increases the probability of ever enrolling in college by 2.5 percentage points (sample means are 11.97, 70 percent, and 41 percent, respectively). Using the triple-difference approach, she finds the same change in the EITC maximum increases the probability of college enrollment by 0.7 percentage points and increases the likelihood of completing a bachelor's degree by 0.3 percentage points (sample means are 26 percent and 3 percent, respectively). Estimated effects are larger for girls and black children. She finds much larger effects for children who were less than 12 at the time of the state EITC implementation, with no effects for children who were college aged.

Milligan and Stabile (2011) examine the effects of the expansion of child tax benefits on child development in Canada using the National Longitudinal Study of Children and Youth and the Survey of Labour and Income Dynamics (SLID). They study two main policies – the Canada Child Tax Benefit and the National Child Benefit program. These programs provide cash assistance based on the number of children and are phased out after a certain level of income. These programs differ significantly from the EITC in two ways. They do not require the parents to work to receive the benefits, and the programs cover a much larger proportion of the population (85 percent of the sample from the SLID receive the Federal benefits). Using a simulated benefits instrumental variables approach, which exploits variation across time, province, and family size, the authors find that increased benefit levels increase achievement. A $1,000 (2004 dollars) increase in benefits increases math scores by 0.069 SD for children ages 6 to 10 and increases vocabulary test scores by 0.149 SD (though not statistically significant) for children ages 4 to 6, with much larger effects for boys on both measures. They also find the tax benefits decrease child aggression and hunger, and reduce maternal depression.

The Welfare-to-Work (WTW) experiments in the 1990s were designed to increase employment and reduce welfare receipt with two main types of programs. The first encouraged work by providing earnings supplements and the second through mandatory employment services and time limits on welfare receipt. The literature generally suggests that programs designed to increase both employment as well as income through income supplements improve child outcomes, while those without income supplements do not have much impact. Existing research only finds evidence of improved outcomes for very young or elementary-aged children, with no positive impacts on adolescents (Morris et al. 2001 and Smolensky and Gootman 2003). Pooling achievement reports across 13 WTW programs, Morris et al. (2005) find that assignment to a WTW program with an earnings supplement increases achievement for children ages 2 to 3 by about 0.070 SD, increases achievement for children ages 4 to 5 by 0.100 SD, and actually decreases achievement for children ages 10-11 by 0.112 SD, with no effects for other programs or ages. For reference, these WTW experiments increased total annual income, which includes earnings, earnings supplements, and AFDC and food stamp benefits, by about $1,750 (2001 dollars).

1.4 Data

I use the restricted geocode data from the NLSY 1979 cohort and the corresponding child file. This data set is a sample of 12,686 young men and women who were age 14 to 20 on December 31, 1978. Individuals are surveyed annually through 1994 and every other year thereafter. Beginning in 1986, children of the mothers in the NLSY are also interviewed every other year. After 1994, children of the NLSY over age 15 are no longer assessed as children and are given a "young adult" survey with questions similar to those asked of the mothers. The NLSY contains extensive information on both the mothers and children, including information on family income and labor market participation and multiple child achievement assessments.

The longitudinal nature of the data allows for direct estimation of long-term effects of EITC expansions on child outcomes that is not possible using a repeated cross section. From the NLSY, I know which state a child lives in as well as family size and income measures each survey year. This allows me to determine state and Federal EITC parameters as well as changes in welfare generosity throughout childhood. Using the ninth version of the National Bureau of

Economic Research's TAXSIM program (Feenberg and Coutts 1993),[2] I am able to estimate a family's tax liability each year, including its state and Federal EITC eligibility and payments. Using these estimates along with welfare receipt reported in the NLSY, I can calculate changes to home resources following EITC expansions, providing some insight into the pathways through which the EITC affects children. Also, the NLSY allows for estimation of models with family fixed effects, as all children in surveyed families are assessed.

I use data on children linked to their mothers for all available years from 1988 through 2000, covering all major Federal expansions of the EITC. The young adult survey provides long-term outcomes for the children that span from 1994 through 2010. Following Dahl and Lochner (2012), I do not include families with mothers who are in the military, in school full-time, or disabled,[3] as these women could have much different labor supply responses to tax changes than other women. To target those families who are actually affected by changes to the EITC, I include those children in the analyses whose family income *ever* fell into the range where they would be eligible to receive the EITC in a given year.[4] I also only include those children who have a sibling in the estimation sample since my preferred estimates include family fixed effects.[5] This sample contains 14,607 child-year observations, with 3,720 children born to 1,424 mothers.

I analyze the effects of the EITC on contemporaneous child achievement and long-term educational attainment. To measure achievement, I use the Peabody Individual Achievement

[2] The program can be accessed at http://nber.org/taxsim.
[3] This restriction eliminates about 3 percent of child-year observations.
[4] Taxable income isn't explicitly given in the NLSY, so I use family earned income (from salary, wages, and tips) to estimate a family's tax liability. Earned income may underestimate taxable income, but the two measures are likely very close for low income families.
[5] Including children without siblings in the analysis would attenuate the long-run results since the outcomes do not vary for each child. About 83% of the "ever-EITC-eligible" children have a sibling in the sample. This restriction results in a slightly more disadvantaged sample. Summary statistics and regression results for the full "ever-EITC-eligible" sample are available upon request.

Test (PIAT) in math and reading comprehension.[6] The math test measures achievement in mathematics as taught in mainstream education, and the reading test measures a child's ability to derive meaning from sentences read silently. The PIAT is one of the most widely used assessments in child achievement research with demonstrated reliability and concurrent validity (Center for Human Resource Research 2004). The tests are administered to children age 5 and older and are normed by age to have a national mean of zero and SD of one. Long-term outcomes include whether a child has a high school diploma or GED, whether he or she has completed one or more years of college, and highest grade completed at age 19.[7] These educational attainment measures are fairly standard in the literature, making comparison to previous studies straightforward.

Table 3 contains summary statistics for this "ever-EITC-eligible" sample of children. About 39 percent of the children are black and 23 percent Hispanic.[8] The average real earned income is $26,332 (2008 dollars) and 41 percent of the sample falls below the poverty line. The average real maximum combined state and Federal EITC value is $2,855 and the average estimated EITC receipt (using TAXSIM) is $929, with receipt ranging from $0 to $7,052. There are about 2.85 children in each family, with the average age of the mother at birth being just over 24. About 31 percent of the children in this sample have mothers with less than a high school education. Child achievement scores are below the national average for PIAT math and reading comprehension at -0.20 and -0.12, respectively (The means in the full NLSY sample are -0.04 and 0.05, respectively). At age nineteen, 75 percent of the children have a high school diploma or

[6] There is also a PIAT in reading recognition that I do not examine because it initially had issues that invalidated scores for young children (Center for Human Resource Research 2004).
[7] As the children of the NLSY are only interviewed every other year, these long-term variables are actually measured when a child is either 19 or 20 in order to include all children in the analysis.
[8] The NLSY oversamples poor black and Hispanic households.

GED, 25 percent have completed one or more years of college, and the average highest grade completed is 12.07 (Full-sample means are 0.81, 0.32, and 12.28, respectively).

TABLE 3
Summary Statistics, 1988-2000

VARIABLES	Obs.	Mean	Std. Dev.	Min	Max
PIAT Math	9908	-0.20	0.88	-2.33	2.33
PIAT Reading Comprehension	8210	-0.12	0.91	-2.33	2.33
HS Diploma or GED (at age 19)	8316	0.75	0.43	0	1
Completed One or More Years College (at age 19)	6382	0.25	0.43	0	1
Highest Grade Completed (at age 19)	7977	12.07	1.23	0	16.00
Age	14607	7.65	3.90	0	14.92
Hispanic	14607	0.23	0.42	0	1
Black	14607	0.39	0.49	0	1
Male	14607	0.50	0.50	0	1
Birth Order	14607	2.19	1.16	1	10
Mother Age at Birth	14607	24.40	4.46	13	41
Mother Married	14607	0.49	0.50	0	1
Mother AFQT Score	14096	29.40	24.03	0	99.49
Mother has less than HS Education	14607	0.31	0.46	0	1
Number of Children in Family	14607	2.85	1.19	1	9
EITC Eligible	14607	0.57	0.49	0	1
Real EITC Maximum ($1000s)	14607	2.85	1.43	1.58	7.40
EITC Payment ($1000s)	14607	0.93	1.27	0	7.05
Real Maximum AFDC Family of 3 ($1000s)	14607	6.94	3.13	1.93	16.90
Any Time Limits on AFDC Receipt	14607	0.17	0.37	0	1
Real K-12 Per Pupil Spending	14607	7.81	1.74	4.36	13.76
Mother in Labor Force	13507	0.68	0.47	0	1
Real Family Earned Income ($1000s)	14607	26.33	30.61	0	637.94
In Poverty	13202	0.41	0.49	0	1
Real AFDC Receipt	14525	1.56	3.53	0	21.85
Real Food Stamp Receipt	14462	1.41	2.32	0	17.06

Notes: Summary statistics of children in the NLSY whose estimated family income ever fell into the EITC-eligible range and who have a sibling in the sample.

1.5 Methodology

EITC receipt depends on earned income, state, year, and number of children. As family income is likely correlated with unobservables that affect maternal labor supply and child outcomes, directly estimating the effect of the amount of EITC receipt will yield biased results. Thus, I exploit exogenous variation in EITC generosity across time, number of children, and state resulting from Federal policy changes and the timing of state adoption of their own EITCs. EITC generosity, as measured by the maximum possible credit a family is eligible for, is generally increasing over time (but not linearly), and the variation across state and number of children can be quite large as discussed in Section 1.2.

I estimate the following model:

$$\begin{aligned} y_{ijst} = \alpha &+ MaxEITC_{jst}\beta_1 + TwoChildren_{jst}\beta_2 + ThreePlusChildren_{jst}\beta_3 \\ &+ \boldsymbol{Welfare_{st}\beta_4} + PPE_{st}\beta_5 + \boldsymbol{X_{ijst}\beta_6} + \delta_t + \gamma_s + \theta_j + \varepsilon_{ijst} \end{aligned} \quad (1)$$

where i indexes child, j indexes mother (family), s indexes state, t indexes year, and ε_{ijst} is an idiosyncratic error term. y_{ijst}, the outcome of interest, can be either a contemporaneous or long-run outcome. $\boldsymbol{X_{ijst}}$ is a row vector of controls including age of the child and its square, mother's score on the Armed Forces Qualification Test (AFQT), indicators for race, sex, interview month, birth order, and birth year of the child, mother's age and its square, and indicators for mother's marital status including whether she was recently married or divorced since the last survey, age at the birth of the child, and highest grade completed.[9] For regressions with long-run outcomes, I

[9] Less than five percent of observations had missing data for mother AFQT score, mother's highest grade completed, or the child's interview month. For these variables I include an indicator for missing values in the regressions. For

also include the child's age in months and its square when the long-run outcome was measured as well as an indicator for the year you would expect the child to graduate high school based on his or her birth month and year.

$MaxEITC_{jst}$ is the maximum EITC credit possible for family j in state s and year t and varies by state, time, and number of children. $TwoChildren_{jst}$ and $ThreePlusChildren_{jst}$ are indicators for how many children are in family j in year t (one child is the omitted group). I also include state and year fixed effects (γ_s and δ_t). Standard errors are clustered at the state level in all regressions. My identification, therefore, is similar to a DiD specification comparing children in states with their own EITCs to those in states without EITCs before and after implementation as well as comparing children in families with 2 or more children to those in families with one child before and after OBRA 1993.

Between 1993 and 1996, 43 states received waivers to experiment with changes to the welfare program, Aid to Families with Dependent Children (AFDC). These waivers generally required work, set time limits for assistance, or increased work incentives (Meyer and Rosenbaum 2000). In 1996, AFDC was replaced with Temporary Assistance for Needy Families (TANF), which also increased the emphasis on work as well as gave states greater discretion in designing their programs (Rowe 2000). To address these changes in welfare policy over the period, I include the vector $\boldsymbol{Welfare_{st}}$, which contains the maximum welfare benefit in state s in year t for a family of three as well as an indicator for whether any time limits or work requirements for welfare receipt had been put in place.[10] I also include PPE_{st}, the real combined

AFQT score, the missing value is replaced as the mean value for AFQT. Since the other variables are entered as dummy variables in the regressions, the missing values are grouped into the same dummy variable.

[10] I obtained the welfare variables from both the Urban Institute's Welfare Rules Database (http://anfdata.urban.org/wrd/WRDWelcome.cfm) and from data used in Meyer and Rosenbaum (2001) that Bruce D. Meyer generously provided.

state and federal current per pupil spending on K-12 public education in state s in year t, to control for changes in government education spending during this period.[11]

As the NLSY follows a sample of women who were ages 14 to 20 at the end of 1978 and their children beginning in 1986, the age distribution of the mothers and children will change systematically over time. I control flexibly for a rich set of characteristics including age of the mother and child as well as year and state dummy variables to remove aggregate time and state effects, but other unobservable characteristics could also be changing in a way that confounds with the timing of changes in $MaxEITC_{jst}$. For example, if a mother has a second child after 1993, the maximum EITC variable increases. However, it could be the case that mothers with more desirable unobservable characteristics have children later in the sample. Therefore, $MaxEITC_{jst}$ could be picking up these differences in unobservables that affect timing of births in the NLSY rather than the actual effect of the policy.[12] To address this, I also estimate the model using family fixed effects, θ_j, which controls for constant unobservable differences across families. In the context of family fixed effects, only cross-time variation in EITC generosity within a family identifies the effect of the policy.

1.6 Results

I first estimate the effect of EITC generosity on contemporaneous child achievement. As the EITC is typically received through a family's tax return in February or March of the next calendar year, I use the EITC maximum from the previous calendar year as the

[11] I obtained the per pupil spending variable from the National Center for Education Statistics' Common Core of Data (http://nces.ed.gov/ccd/).
[12] This discussion is abstracting from the possibility that families might react to 1993 OBRA by having a second child in order to receive a higher EITC payment. I ignore this, as previous work finds no effect of EITC changes on childbearing (Baughman and Dickert-Conlin 2003 and 2009 and Hotz and Scholz 2003).

“contemporaneous” measure compared to the current year’s test scores. Table 4 contains tabulations for the interview month of the child in the NLSY, which is when he or she takes the PIAT. 99.99 percent of the children are interviewed in April or later and 92.77% in June or later. Therefore, the results should reflect the effects of any changes to maternal labor supply and earnings induced by a change in the maximum value of the EITC in the previous calendar year as well as any immediate effects of the increase in the lump sum EITC payment received with the tax return in the current year.

TABLE 4
Tabulation of Interview Month of Child, 1988-2000

Interview Month	Obs.	Percent	Cumulative Percent
January	2	0.01	0.01
February	0	0	0.01
March	0	0	0.01
April	148	1.01	1.03
May	906	6.2	7.23
June	1,835	12.56	19.79
July	3,236	22.15	41.95
August	4,156	28.45	70.4
September	2,405	16.46	86.86
October	1,073	7.35	94.21
November	462	3.16	97.37
December	107	0.73	98.1
Missing	277	1.9	100
Total	14,607	100	-

Notes: Tabulations for children in the NLSY whose estimated family income ever fell into the EITC-eligible range and who have a sibling in the sample.

Table 5 presents the main Ordinary Least Squares (OLS) results from equation (1) *without* family fixed effects for both the contemporaneous achievement and long-run educational

TABLE 5
Ordinary Least Squares Results, 1988-2000

	Contemporaneous		Long-Run		
VARIABLES	Math	Reading	High School Diploma or GED	Completed 1 or More Yrs. College	Highest Grade Completed
MaxEITC	0.0352	0.0651*	0.0541***	0.0616**	0.0474
	(0.0348)	(0.0364)	(0.0198)	(0.0255)	(0.0623)
Married	0.0141	0.0120	0.0464**	0.0377*	0.1420***
	(0.0290)	(0.0339)	(0.0198)	(0.0200)	(0.0506)
Two Children	0.0064	-0.0534	-0.0192	-0.0315	0.2150**
	(0.0687)	(0.0855)	(0.0359)	(0.0352)	(0.0822)
Three Plus Children	0.0263	0.0157	-0.0011	0.0107	0.0373*
	(0.0175)	(0.0236)	(0.0066)	(0.0067)	(0.0216)
Welfare Max Benefit	-0.0206	-0.0872	-0.0266	-0.0361	0.0632
	(0.0581)	(0.0728)	(0.0350)	(0.0407)	(0.0826)
Time Limits on Welfare	-0.0186	-0.0097	0.0112	-0.0315	-0.0900
	(0.0472)	(0.0368)	(0.0126)	(0.0280)	(0.0566)
PPE	-0.0188	0.0016	0.0029	-0.0208	-0.1010**
	(0.0289)	(0.0420)	(0.0119)	(0.0170)	(0.0419)
Age	0.0492	-0.6050***	-0.0665	-0.0592	-0.5390***
	(0.0707)	(0.0666)	(0.0472)	(0.0492)	(0.119)
Age^2	-0.0081***	0.0142***	-0.0000	0.0000	0.0001
	(0.0010)	(0.0016)	(0.0002)	(0.0002)	(0.0004)
Male	-0.0434	-0.1420***	-0.0960***	-0.1060***	-0.3130***
	(0.0318)	(0.0290)	(0.0242)	(0.0223)	(0.0665)
Hispanic	-0.1850***	-0.0657	0.0679***	0.0434	0.1450**
	(0.0458)	(0.0557)	(0.0247)	(0.0268)	(0.0624)
Black	-0.1630***	0.0132	-0.0194	-0.0592**	-0.0799
	(0.0393)	(0.0373)	(0.0294)	(0.0244)	(0.0781)
Mother AFQT	0.0092***	0.0104***	0.0015***	0.0004	0.0045***
	(0.0009)	(0.0010)	(0.0005)	(0.0006)	(0.0015)
Year Fixed Effects	x	x	x	x	x
State Fixed Effects	x	x	x	x	x
Family Fixed Effects	-	-	-	-	-
Observations	9,808	8,128	8,220	6,310	7,896
R-squared	0.182	0.280	0.192	0.244	0.243

Notes: Robust standard errors (clustered at the state level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Each column represents a separate OLS regression for the various outcomes. Full set of controls also include indicators for interview month, birth order, and birth year of the child, mother's age and its square, and indicators for whether mother was recently married or divorced since the last survey, mother's age at the birth of the child, and mother's highest grade completed. For regressions with long-run outcomes, additional controls include the child's age in months and its square when the long-run outcome was measured as well as an indicator for expected high school graduation year.

attainment measures for the “ever-EITC-eligible” sample.[13] The $MaxEITC_{jst}$ variable is in thousands of real 2008 dollars. An increase in the maximum possible EITC a family can receive in a given year of $1,000 leads to an increase in math scores of 0.035 SD and increase in reading scores by 0.065 SD, with only the reading results being statistically significant at the 10 percent level. A $1,000 increase in $MaxEITC_{jst}$ in a single year increases the probability of receiving a high school diploma or GED at age 19 by 5.4 percentage points (sample mean 75 percent) and the probability of completion of one or more years of college at age 19 by 6.2 percentage points (sample mean of 25 percent), both statistically significant at the 5 percent level or lower. Though not significant, I find that highest grade completed increases by .047 (sample mean of 12.07).

Table 6 presents the analogous results with the inclusion of family fixed effects. I find that a $1,000 increase in $MaxEITC_{jst}$ in a single year increases math achievement by 0.072 SD and reading achievement by 0.039 SD, with the math result being statistically significant at the 5 percent level. The same increase in $MaxEITC_{jst}$ increases the probability of high school diploma or GED receipt by 2.1 percentage points and probability of completion of one or more years of college at age 19 by 1.4 percentage points (significant at the 5 and 10 percent level, respectively). I estimate a positive but statistically insignificant effect on highest grade completed of 0.030.

These estimates are smaller in magnitude than the estimates without family fixed effects with the exception of that for math. The change in the estimates suggests that fixed, unobservable characteristics of families are positively correlated with the maximum EITC value.

[13] For brevity, not all regression coefficients on control variables are shown in the tables, but these regressions contain the full set of controls above.

As described above, a possible explanation for this is that the timing of births in the NLSY is endogenous due to the cohort nature of the survey. Children born later in the sample have higher

TABLE 6
Ordinary Least Squares Results w/ Family Fixed Effects, 1988-2000

	Contemporaneous		Long-Run		
VARIABLES	Math	Reading	High School Diploma or GED	Completed 1 or More Yrs. College	Highest Grade Completed
MaxEITC	0.0717**	0.0388	0.0207**	0.0139*	0.0295
	(0.0274)	(0.0426)	(0.0099)	(0.0078)	(0.0301)
Married	0.0717*	0.0247	0.0067	0.0027	0.0177
	(0.0381)	(0.0388)	(0.0061)	(0.0041)	(0.0140)
Two Children	-0.0285	-0.0520	-0.0133	-0.0154	0.0312
	(0.0568)	(0.0707)	(0.0113)	(0.0122)	(0.0359)
Three Plus Children	-0.0318	0.0045	-0.0074	-0.0124	0.0298
	(0.0528)	(0.0737)	(0.0108)	(0.0099)	(0.0404)
Welfare Max Benefit	0.0478***	0.0046	-0.0014	-0.0019	0.0117
	(0.0162)	(0.0205)	(0.0030)	(0.0033)	(0.0143)
Time Limits on Welfare	0.0263	-0.0294	-0.0055	-0.0041	-0.0142
	(0.0417)	(0.0283)	(0.0105)	(0.0081)	(0.0238)
PPE	0.0194	0.0057	0.0101	0.0099*	-0.0246
	(0.0355)	(0.0514)	(0.0072)	(0.0057)	(0.0326)
Age	0.1810**	-0.3570***	-0.1070*	0.0099	-0.3500*
	(0.0769)	(0.0740)	(0.0605)	(0.0753)	(0.1880)
Age^2	-0.0087***	0.0108***	-0.0000	-0.0000	-0.0004
	(0.0011)	(0.0018)	(0.0001)	(0.0001)	(0.0003)
Male	-0.0086	-0.1250***	-0.0838***	-0.1100***	-0.3570***
	(0.0283)	(0.0317)	(0.0271)	(0.0290)	(0.0942)
Year Fixed Effects	x	x	x	x	x
State Fixed Effects	x	x	x	x	x
Family Fixed Effects	x	x	x	x	x
Observations	9,808	8,128	8,220	6,310	7,896
R-squared	0.493	0.591	0.730	0.809	0.738

Notes: Robust standard errors (clustered at the state level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Each column represents a separate OLS regression for the various outcomes. Full set of controls also include indicators for interview month, birth order, and birth year of the child, mother's age and its square, and indicators for whether mother was recently married or divorced since the last survey, mother's age at the birth of the child, and mother's highest grade completed. For regressions with long-run outcomes, additional controls include the child's age in months and its square when the long-run outcome was measured as well as an indicator for expected high school graduation year.

possible EITC payments, and the results suggest that they are also born to families with more desirable fixed unobservable characteristics. I prefer these estimates for this reason and include family fixed effects for the remaining analyses.

1.6.1 Heterogeneity in the Results

The above analysis assumes that the effects of an increase in EITC generosity in a given year has the same effect on the both the contemporaneous and long-run child outcomes for all children in each year of his or her childhood. However, previous EITC, income, and maternal labor supply studies find heterogeneity across subgroups, particularly by age and sex of the child. Table 7 presents results for various subgroups of the sample. These estimates are similar to above with family fixed effects, but with the $MaxEITC_{jst}$ variable interacted with indicators for the subgroups. I also include indicators for subgroup separately if this varies within family. For example, when looking separately by sex of the child, $MaxEITC_{jst}$ would be replaced with $MaxEITC_{jst} * male_i$, $MaxEITC_{jst} * female_i$, and $male_i$ (where $male_i$ and $female_i$ are dummy variables).

The first row of Table 7 reproduces the results on the full sample from Table 6. I first look at effects by age of the child, where I define preschool age as less than 4 years old, elementary age as between 4 and 11, and middle school age as between 11 and 15. Consistent with Chetty et al. (2011a), I find larger effects on contemporaneous achievement for middle school aged children compared to elementary school for both math and reading, but the estimates aren't statistically different (0.075 SD versus 0.067 SD for math and 0.045 SD versus 0.025 SD for reading, respectively).

TABLE 7
Ordinary Least Squares Results with Family Fixed Effects by Subgroups, 1988-2000

	Contemporaneous		Long-Run		
VARIABLES	Math	Reading	High School Diploma or GED	Completed 1 or More Yrs. College	Highest Grade Completed
All	0.0717**	0.0388	0.0207**	0.0139*	0.0295
	(0.0274)	(0.0426)	(0.0099)	(0.0078)	(0.0301)
Preschool	-	-	0.0359	0.0259	0.1110
			(0.0334)	(0.0442)	(0.0798)
Elementary	0.0673**	0.0250	0.0240**	0.0161	0.0323
	(0.0334)	(0.0454)	(0.0117)	(0.0102)	(0.0371)
Middle School	0.0745***	0.0453	0.0193*	0.0132*	0.0304
	(0.0259)	(0.0421)	(0.0096)	(0.0076)	(0.0279)
Boys	0.0934***	0.0500	0.0220**	0.0140*	0.0366
	(0.0293)	(0.0444)	(0.0103)	(0.0080)	(0.0294)
Girls	0.0474	0.0235	0.0182*	0.0138*	0.0220
	(0.0307)	(0.0413)	(0.0100)	(0.0081)	(0.0323)
Minority	0.0894***	0.0210	0.0232**	0.0138*	0.0302
	(0.0281)	(0.0444)	(0.0100)	(0.0079)	(0.0299)
White	0.0593**	0.0536	0.0183*	0.0140*	0.0286
	(0.0265)	(0.0431)	(0.0099)	(0.0080)	(0.0308)

Notes: Robust standard errors (clustered at the state level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
OLS regression for various subgroups with full set of controls from Table 6.

Though not statistically different from one another, the magnitudes on the estimates for all long-run outcomes monotonically decrease as the age band increases. For example, I estimate that a $1,000 increase in $MaxEITC_{jst}$ during preschool increases the probability of high school or GED completion by about 3.6 percentage points. This same increase during middle school increases this probability by only about 1.9 percentage points. One possible explanation is that a child who is young during an EITC expansion likely receives higher EITC payments for the remainder of his or her childhood as well, whereas an older child would only benefit from the more generous EITC for a few years. Michelmore (2013) also finds larger effects of state EITCs for children who were less than 12 years old at the time of the state EITC adoption. Another possibility is that developmental malleability is much stronger for very young children (Shonkoff and Phillips 2000). Duncan et al. (1998) find that family economic circumstances before age five are more predictive of children's completed schooling than at ages 6 to 15.

Looking separately by sex of the child, I find much larger effects for boys compared to girls on all outcome measures. For math achievement, the estimated effect for boys is almost twice as large as that for girls (0.093 SD and 0.047 SD, respectively), though they are not statistically different at conventional levels. This is consistent with previous studies finding much larger effects of income via tax credits on achievement for boys (Milligan and Stabile 2011 and Dahl and Lochner 2012). I also estimate larger effects for boys on all long-run outcomes though, again, the estimates aren't statistically different than the estimates for girls.

Lastly, I estimate effects separately by race. Again consistent with Dahl and Lochner (2012), I find larger effects on math achievement for minority children (black or Hispanic) compared to their white counterparts (0.089 SD and 0.059 SD, respectively). Estimates on long-run outcomes are fairly similar for the two groups, but I estimate a larger effect for minority children on high school diploma or GED receipt (2.3 and 1.8 percentage points, respectively). Michelmore (2013) also finds larger effects of the EITC on long-term educational attainment for minority children. In the "ever-EITC-eligible" sample, average real earned income is about $21,500 for minority families and about $34,200 for white families. As minority status is a crude proxy for income, this finding suggests that the EITC is more effective at improving educational outcomes for the most disadvantaged children.

1.6.2 Interpreting the Magnitudes of the Effects

The above coefficient estimates represent the effects of a $1,000 increase in the maximum EITC benefit a family is eligible for in a given year. To interpret the estimates, it is helpful to determine how this change in EITC generosity affects maternal labor supply and family income. Table 8 presents results using the same sample and methodology as above (including family fixed effects) where I regress various measures of family income and maternal labor supply on the maximum EITC variable. For the "ever-EITC-eligible" sample, a $1,000 increase in $MaxEITC_{jst}$ increases maternal labor force participation by about 6.4 percentage points and increases yearly hours worked by about 93.3. These results are consistent with previous EITC maternal LFP findings (Dickert et al. 1995; Eissa and Liebman 1996; Ellwood 2000; Meyer and Rosenbaum 2000 and 2001; Neumark and Wascher 2001; Grogger 2003; Eissa and Hoynes 2004; Hotz et al. 2006; Rothstein 2007; and Adireksombat 2010) as well as the labor supply incentives created by the EITC.[14]

Using the NBER's TAXSIM program and reported earnings from the NLSY, I estimate each family's tax liability and EITC payment. A $1,000 increase in $MaxEITC_{jst}$ increases average estimated EITC receipt by about $328, increases average estimated after-tax income (including EITC) by about $1,446, and decreases average tax liability (including the EITC) by about $598 in the sample. A $1,000 increase in the EITC maximum reduces AFDC/TANF

[14] Running the maternal labor force participation regressions separately by marital status yields point estimates of 0.142 for families with single mothers and -0.026 for those with married mothers.

TABLE 8
OLS Results with Family Fixed Effects on Maternal LS and Family Income, 1988-2000

Mother In LF	*Hours Worked*	*EITC Payment*	*After-Tax Income*
0.0640***	93.25*	0.328***	1.446
(0.0175)	(55.56)	(0.108)	(1.182)
Tax Liability	*AFDC/TANF*	*Food Stamps*	*Total Net Income*
-0.598	-0.525*	-0.135**	0.888
(0.388)	(0.276)	(0.065)	(1.110)

Notes: Robust standard errors (clustered at the state level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
OLS regression for various outcomes with full set of controls from Table 6.

receipt by about $525 and reduces food stamp receipt by about $135 (both AFDC/TANF and food stamp receipt are reported in the NLSY).

I create a net income measure using reported earned income and welfare receipt and estimated tax liability/EITC. I estimate that family net income increases by about $888 following an increase in the EITC maximum of $1,000. I therefore interpret my estimates as the effect of a net increase in income of about $888. Using this interpretation, my estimate of a 0.072 SD increase in math is very comparable in magnitude to the other EITC studies. Dahl and Lochner (2012) find that a $1,000 (2000 dollars) increase in income as a result of EITC expansions increases math achievement by 0.058 SD, and Chetty et al. (2011a) find that a $1,000 (2010 dollars) increase in EITC receipt holding earned income constant increases math achievement by 0.093 SD. Both of these studies find larger effects on reading than my estimate suggests. I estimate a 0.039 SD increase in reading, and Dahl and Lochner and Chetty et al. estimate effects of 0.061 SD and 0.062, respectively. However, note that my estimated effect for reading isn't precisely measured.

For illustration on the economic importance of the effects, consider an elementary school aged child in a family of two after OBRA 1993 is fully phased in. In the absence of a state EITC, this child would be eligible for a maximum credit of about $5,000. Compared to the maximum credit of about $2,000 before OBRA 1993, my fixed effects estimates from Table 7 suggest that this child would have a higher math score of about 0.215 SD, an increased probability of graduating high school or receiving a GED of about 7.2 percentage points (9.4% increase from sample mean), and an increased probability of completing one or more years of college by age 19 by about 4.8 percentage points (18.5% increase).

For comparison, consider one of the most studied education experiments, the Student/Teacher Achievement Ratio (STAR) experiment in Tennessee in the 1980s aimed to determine the effects of class size in kindergarten through third grade. Krueger (1999) and Chetty et al. (2011c) find that students assigned to a small class in kindergarten (about 15 students compared to 23 students) score about 4 percentile points, or about 0.20 SD, higher on combined math and reading achievement that year. On average, students assigned to a small class spend 2.14 years longer in a small class than those assigned to a large class. Chetty et al. (2011c) find that students assigned to a small class are 1.8 percentage points more likely to attend college at age 20, a 26.4% increase in their sample. Using a comparison of means of the STAR data, Finn et al. (2004) find that four years in a small class is associated with a significantly higher graduation rate than attending full-size classes (87.8% and 76.3%, respectively, suggesting a 14% increase from the sample mean).

These studies indicate that my estimated effects of OBRA 1993 on achievement and educational attainment are fairly comparable in magnitudes to those found from STAR. For comparison, consider the costs of the two programs. Krueger and Whitmore (2001) estimate the cost of the STAR program using average U.S. per pupil spending data. A treated student spends 2.3 years in a small class which amount to a cost of about \$10,712 (2008 dollars). My estimates above for OBRA 1993 are for effects on elementary aged children. The average age of elementary children in my sample is 7.56, indicating that a child will receive this higher EITC for about 11.44 years until the child is 19. This amounts to an \$11,257 increase in EITC spending and a total cost to the government, the change in tax liability over time, of \$20,180. However, this cost does not take into account the changes in government spending on other welfare programs following an EITC expansion. My estimates suggest that the change in tax liability is actually smaller than the decrease in combined AFDC/TANF and food stamp receipt, indicating that this program might actually have an overall *negative* cost to the government (both state and Federal combined).

Another important input for educational outcomes is teacher quality. Rockoff (2004), Rivkin, Hanushek, and Kain (2005), and Kane and Staiger (2008) estimate that a 1 SD increase in teacher quality raises test scores by between 0.1 and 0.2 standard deviations. Thus, my estimates suggest that OBRA 1993 (\$3,000 increase in EITC maximum) had a similar impact on test scores for elementary and middle school aged children as a 1 SD increase in teacher quality. Chetty et al. (2011a) find that a 0.2 SD increase in test scores in a single grade as a result of being assigned a higher quality teacher raises the probability of college attendance at age 20 by about 1.0 percentage points (sample mean of 37%). I estimate that OBRA 1993 had a similar effect on test scores, but much larger long-term gains on college attendance. I estimate that OBRA 1993 increased the probability of completing one or more years of college at age 19 by 4.8 percentage points for children in elementary school during the law change and 4.0 percentage points for children in middle school (sample mean of 25%).

1.6.3 Specification Checks

I check the robustness of my results to alternative specifications in the top panel of Table 9. The first line again reproduces my main results including family fixed effects from Table 6. I first estimate the model using the natural log of the maximum EITC variable. I find no difference in the patterns of the results, but have less power in identifying effects. I next estimate the model using the NLSY-provided sample weights. These weights are designed to correct for the over-sampling of low income black and Hispanic households, yielding a nationally representative sample each year of children born to mothers age 14 to 20 at the end of 1978. However, when selecting the sample using variables with missing values (in this case, earnings), the weights don't yield this nationally representative sample. Generally, using the weights provides a noisier estimate that more heavily weights the observations of white children in the sample. Using the weights, I find larger effects for reading and highest grade completed and smaller effects for the other outcomes. Lastly, I estimate the model using only the Federal maximum value of the EITC. These results are not statistically different from the original specification.

The bottom panel of Table 9 contains results for three falsification tests. The first line of estimates is that from a test in which I estimate the specification from equation (1) on the various outcomes, but on the sample of children whose families were never in the EITC-eligible range during this time period.[15] As these children never received the EITC, they should not be affected by changes in its generosity over time. Finding an effect in this sample could indicate that my identification strategy is falsely attributing the effects of shocks that impact all children over time to changes in the maximum EITC. All estimates on the $MaxEITC_{st}$ variable for this "never-EITC-eligible" sample are statistically insignificant with the exception of the college completion

[15] I again include only those children with a sibling in the estimation sample.

estimate. This estimate is statistically significant at the 10 percent level, but the point estimate is negative. All estimates are also smaller in absolute magnitude than the original estimates except for reading. I estimate a larger, but very imprecisely measured effect on reading in this sample.

TABLE 9
Robustness to Alternative Specifications and Falsification Tests, 1988-2000

	Contemporaneous		Long-Run		
VARIABLES	Math	Reading	High School Diploma or GED	Completed 1 or More Yrs. College	Highest Grade Completed
Original	0.0717**	0.0388	0.0207**	0.0139*	0.0295
	(0.0274)	(0.0426)	(0.0099)	(0.0078)	(0.0301)
Log MaxEITC	0.2650*	0.2620	0.0522	0.0422	0.1140
	(0.1360)	(0.1860)	(0.0407)	(0.0297)	(0.1250)
Weighted	0.0315	0.0177	0.0196**	0.0052	0.0475
	(0.0290)	(0.0512)	(0.0089)	(0.0066)	(0.0306)
Only Federal MaxEITC	0.1140***	0.0163	0.0214	0.0118	0.0154
	(0.0401)	(0.0652)	(0.0143)	(0.0130)	(0.0551)
Non-EITC eligible	0.0062	0.0740	0.0041	-0.0120*	-0.0172
	(0.0465)	(0.0623)	(0.0037)	(0.0070)	(0.0170)
2 vs. 3+ Children	0.0225	0.0040	-0.0007	-0.0009	-0.0009
	(0.0212)	(0.0226)	(0.0047)	(0.0042)	(0.0119)
2 vs. 3+ Children	0.0080	-0.0090	-0.0043	-0.0025	-0.0140
Only Federal EITC	(0.0217)	(0.0222)	(0.0049)	(0.0044)	(0.0134)

Notes: Robust standard errors (clustered at the state level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Each cell represents a separate OLS regression for various outcomes with full set of controls from Table 6.

OBRA 1993 increased the EITC credit differentially for families with one child compared to those with two or more children, but, in all states except Wisconsin, the maximum EITC payment does not differ for families with 2 or more children. In the last two lines of Table

9, I conduct a falsification test for families with 2 or more children where I assign families with 2 children the maximum EITC value for a family with one child and assign the families with 3 or more children their actual EITC maximum values. I also exclude children living in Wisconsin from the estimation. Without state EITCs, this test basically amounts to a difference-in-differences estimation comparing children in families with 2 children to families with 3 or more children before and after OBRA 1993.

Finding a positive effect on this “false” maximum EITC variable could indicate that my main results are incorrectly attributing effects due to the timing of births in the NLSY as effects of changes in EITC generosity. The first line of results contains state variation in the “false” maximum EITC over time, and the last line contains only Federal variation. In both specifications, the estimates on all outcomes are much smaller in absolute magnitude and negative in most cases, none of which are close to statistical significance. These falsification tests provide strong support for the validity of my research design including family fixed effects.

1.7 Summary and Conclusions

I find that an increase in the generosity of the EITC has large positive impacts on both contemporaneous child achievement and long-run educational attainment. An increase in the maximum EITC of $1,000 in a given year significantly increases math achievement by about 0.072 SD for children in families who were ever eligible for EITC receipt. This change in EITC generosity during childhood also significantly increases the probability of receiving a high school diploma or GED at age 19 by about 2.1 percentage points and the probability of completing one or more years of college at age 19 by about 1.4 percentage points. I find larger effects for boys and minority children and evidence that an expansion in the EITC is more effective at improving educational outcomes for children who are younger during the expansion. Along with changes to child educational outcomes, an increase in the maximum EITC of $1,000 results in other changes in the household, including an increase in net family income inclusive of EITC and welfare payments of about $888 and an increase in maternal labor force participation.

Overall, the EITC appears to be an effective policy for improving educational outcomes of children, especially for the most disadvantaged. In the current context of work requirements and lifetime limits for TANF and with recent cuts to the food stamp program, the EITC is likely more important now for low income families than this study suggests. As more data become available from the NLSY or other sources, additional work is needed to determine the effects of the EITC on other long-term outcomes of children such as earnings or welfare dependency in order to fully assess the performance of the program. While the NLSY contains these variables, the children are not yet old enough in the available data to analyze these outcomes.

CHAPTER 2

The Effects of the Earned Income Tax Credit on Net Family Financial Resources

2.1 Introduction

The Earned Income Tax Credit (EITC) is a significant source of government aid to low income families and has grown dramatically since its inception in 1975. Total outlay reached over $50 billion in 2008, more than tripling the spending on government cash welfare assistance via Temporary Assistance to Needy Families (Internal Revenue Service 2011). There are numerous studies examining the effects of the EITC on various measures of parental and child well-being. The program appears to be effective at improving maternal mental and physical health (Evans and Garthwaite 2014), increasing prenatal care, and significantly reducing the likelihood of low birth weight (Baker 2008; Strully et al. 2010; and Hoynes et al. 2012). Expansions in the EITC also improve child test scores (Chetty et al. 2011a; Dahl and Lochner 2012; and Maxfield 2013) and increase educational attainment and college attendance (Maxfield 2013; Michelmore 2013; and Manoli and Turner 2014). Despite these findings and the size of the program, we still do not have a clear understanding of how the EITC affects the overall monetary resources of families receiving the credit.

Previous studies find that EITC expansions raise earnings and bring families above the poverty line (Dahl et al. 2009; Neumark and Wascher 2001; Grogger 2003; Meyer 2007; and Strully et al. 2010). However, these measures do not give a full picture of financial well-being, as they do not take into account all income sources of a family. EITC expansions affect multiple income streams including earnings, welfare and food stamp receipt, and EITC payments. For example, EITC expansions increase labor supply and earnings, which mechanically reduces welfare and food stamp receipt. One could imagine that a change in welfare receipt, for example, is not the same to a family as a change in its EITC payment, which is typically delivered as a lump sum payment once a year. An accurate measure of the effect of the EITC on various income sources and overall net income could provide some insight into the mechanisms through which the EITC affects families.

Using the March Current Population Survey (CPS), I examine the effect of the 1993 expansion of the EITC on various measures of financial well-being of families likely receiving the credit using a difference-in-differences (DiD) framework. I construct a net family income measure that includes nearly all income sources – earned income, unearned income, government assistance, and the EITC and other taxes. This comprehensive of an income measure has not been previously examined in the context of EITC expansions. I also examine the effect of the EITC on the various income sources comprising the net income measure separately. I find significant evidence that the 1993 expansion of the EITC improves the financial well-being of its recipients as measured by the comprehensive income measure. I find that a relative increase in the real maximum EITC of about $1,900 (2008 dollars) increases real net annual family income by about $527 for all low income families coupled with a sizable increase to maternal labor force participation for those families with single mothers. The increase in overall net income is a result of an increase in earnings and EITC payments and a smaller decline in welfare and food stamp receipt.

In the following section, I review the institutional details of the EITC. Section 2.3 outlines how the EITC affects family monetary resources and reviews the previous literature on this topic. Section 2.4 describes the CPS data and presents summary statistics for my sample. Section 2.5 details my empirical strategy, and Section 2.6 presents the results. I summarize the findings and conclude in Section 2.7.

2.2 Institutional Details of the EITC

The EITC began in 1975 with modest credits for low income families with children as a way to offset payroll taxes. Since then, the Federal government expanded the EITC multiple times in an effort to create an anti-welfare, anti-poverty, and pro-work tool (Ventry 2000). The credit is refundable and only available to families who work. It is based on a family's earned income, number of children, and state of residence. Table 10 shows the Federal EITC parameters for the years I examine, 1990 to 1999. As the table illustrates, there is an initial "phase-in" range and rate, where the credit is equal to the subsidy rate times the family's earned income until the maximum credit is reached. The family then receives the maximum credit during the "flat" range. Once a family reaches a certain level of income, they enter a "phase-out" range, where the credit is reduced at the phase-out rate. Thus, only families below a certain level of income are eligible for the credit in each year. Families are given the option to receive the credit with periodic payments throughout the year as opposed to a one-time lump sum. However, less than five percent of families exercised this option during the time frame I study (Friedman 2000). Thus, the vast majority of families receive their EITC credit as a lump sum upon filing their tax returns, with over 80 percent of families receiving the credit by the end of March (LaLumia 2013).

In addition to Federal funding of the credit, states have the option to add their own EITCs that typically "piggyback" onto the Federal credits – meaning these states will increase the Federal EITC credit by a given percentage. The states vary substantially on the generosity of their credit, whether they offer it to families without children, and whether the credit is refundable. Table 11 contains the state EITC parameters from 1990 to 1999. By 1999, eleven states enacted their own EITCs and the state add-ons range from 4 to 75 percent in this period.

TABLE 10
Federal EITC Parameters, 1990 – 1999

Calendar year	Credit rate (%)	Min income for max credit	Max credit	Phase-out rate (%)	Phase-out range: Beginning income	Phase-out range: Ending income
1990	14	6,810	953	10	10,730	20,264
1991						
One child	16.7	7,140	1,192	11.93	11,250	21,250
Two children	17.3	7,140	1,235	12.36	11,250	21,250
1992						
One child	17.6	7,520	1,324	12.57	11,840	22,370
Two children	18.4	7,520	1,384	13.14	11,840	22,370
1993						
One child	18.5	7,750	1,434	13.21	12,200	23,050
Two children	19.5	7,750	1,511	13.93	12,200	23,050
1994						
No children	7.65	4,000	306	7.65	5,000	9,000
One child	26.3	7,750	2,038	15.98	11,000	23,755
Two children	30	8,425	2,528	17.68	11,000	25,296
1995						
No children	7.65	4,100	314	7.65	5,130	9,230
One child	34	6,160	2,094	15.98	11,290	24,396
Two children	36	8,640	3,110	20.22	11,290	26,673
1996						
No children	7.65	4,220	323	7.65	5,280	9,500
One child	34	6,330	2,152	15.98	11,610	25,078
Two children	40	8,890	3,556	21.06	11,610	28,495
1997						
No children	7.65	4,340	332	7.65	5,430	9,770
One child	34	6,500	2,210	15.98	11,930	25,750
Two children	40	9,140	3,656	21.06	11,930	29,290
1998						
No children	7.65	4,460	341	7.65	5,570	10,030
One child	34	6,680	2,271	15.98	12,260	26,473
Two children	40	9,390	3,756	21.06	12,260	30,095
1999						
No children	7.65	4,530	347	7.65	5,670	10,200
One child	34	6,800	2,312	15.98	12,460	26,928
Two children	40	9,540	3,816	21.06	12,460	30,580

Source: Joint Committee on Taxation, Ways and Means Committee (2004).
Note: Dollar amounts unadjusted for inflation

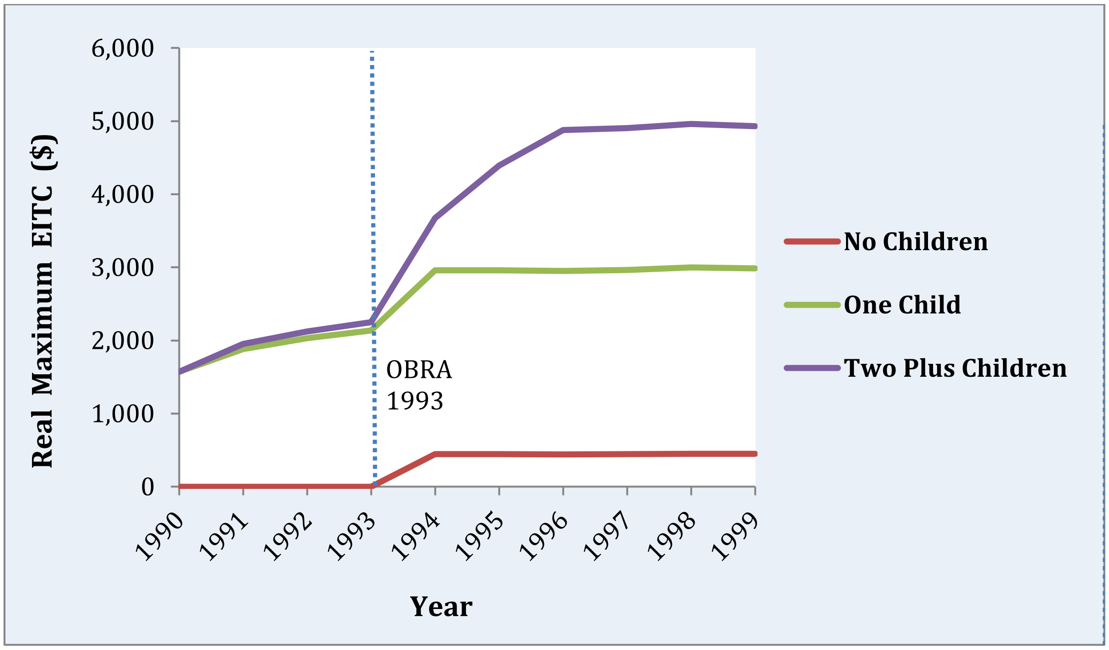
TABLE 11
State EITC Supplements, 1990 – 1999 (%)

State	CO	IA	KS	MA	MD	MD	MN	MN	NY	OR	RI	VT	WI	WI	WI
No. Children	0+	0+	0+	0+	1+	1+	0	1+	0+	0+	0+	0+	1	2	3+
1990		5			50						28	28	5	25	75
1991		6.5			50		10	10			28	28	5	25	75
1992		6.5			50		10	10			28	28	5	25	75
1993		6.5			50		15	15			28	28	5	25	75
1994		6.5			50		15	15	7.5		28	25	4.4	21	63
1995		6.5			50		15	15	10		28	25	4	16	50
1996		6.5			50		15	15	20		28	25	4	14	43
1997		6.5		10	50		15	15	20	5	28	25	4	14	43
1998		6.5	10	10	50	10	15	25	20	5	27	25	4	14	43
1999	8.5	6.5	10	10	50	10	25	25	20	5	27	25	4	14	43
Refundable?	Y	N	Y	Y	N	Y	Y	Y	Y	N	N	Y	Y	Y	Y

Sources: Center on Budget and Policy Priorities and Leigh (2010).
Notes: No. Children is the number of children required for eligibility of the state supplement. Supplement is the percentage top-up of the federal EITC payment.

Figure 4 plots the real (2008 dollars) value of the maximum Federal EITC credit by tax year and number of children. The 1993 enactment of the Omnibus Reconciliation Act (OBRA 1993) resulted in real expansions in the Federal maximum credit. The changes were quite substantial and also increased the Federal maximum EITC differentially by number of children for the first time. For example, families of all sizes would be eligible for about the same real Federal maximum EITC credit of about \$2,200 in 1993. In 1996, after OBRA 1993 was fully phased in, families with one child would be eligible for a maximum credit of about \$3,000, and families with two or more children would be eligible for a much higher maximum credit of about \$4,900. This differential change by family size is the basis of my identification strategy discussed below.

FIGURE 4
Real Maximum Federal EITC Credit by Tax Year and Number of Children (2008$)



Sources: Joint Committee on Taxation, Ways and Means Committee (2004). Formatting adopted from Hoynes et al. (2012).

2.3 The EITC and Family Financial Resources

The EITC affects net family financial resources through changes in the labor supply decisions of mothers and also changes in various categories of family income. The structure of the credit provides incentives for altering child bearing and marriage decisions as well, but previous studies find no effect of the EITC on these outcomes (Eissa and Hoynes 2000; Ellwood 2000; Dickert-Conlin and Houser 2002; Hotz and Scholz 2003; and Baughman and Dickert-Conlin 2003 and 2009). Thus, I focus this discussion on the effects of the EITC on maternal labor supply and family income.

The structure of the EITC creates different labor supply incentives depending upon the taxable income of the family. Assuming leisure is a normal good and the mother is not filing jointly with a spouse with positive earnings (typically a single mother), an EITC expansion creates an unambiguously positive incentive to enter the labor force, as it increases the potential wage of those not participating in the labor force.[16] For those sole-earner mothers already participating, the incentive depends upon income and the EITC parameters in a given year. If the mother is working and her income falls in the "phase-in" range of the EITC, there is a substitution effect away from leisure since the EITC-induced wage increase makes leisure more expensive. There is also an income effect to consume more leisure. Thus, the overall effect on hours worked is ambiguous. By similar reasoning, women in the "flat" or "phase-out" range have an unambiguous incentive to work less. Women with family income above the cutoff to be eligible for the EITC (end of the phase-out range) may also have an incentive to reduce their labor supply depending on their preferences and how close they are to the end of the phase-out range. As the EITC is based on family earnings, mothers filing jointly with a wage-earning husband will be more likely to fall in the flat or phase-out range of the EITC schedule. Therefore, some of these women may be induced to decrease their hours worked, or possibly even leave the labor force altogether (see Hotz and Scholz 2003 for a more detailed theoretical discussion of labor supply responses to the EITC).

Previous work confirms these labor supply predictions. First, EITC expansions substantially increase the labor force participation (LFP) of single mothers (Dickert et al. 1995; Eissa and Liebman 1996; Ellwood 2000; Meyer and Rosenbaum 2000 and 2001; Neumark and Wascher 2001; Grogger 2003; Hotz et al. 2006; Rothstein 2007; and Adireksombat 2010). If anything, the credit *modestly* decreases the LFP of married mothers (Dickert et al. 1995; Ellwood 2000; and Eissa and Hoynes 2004).

[16] Technically you must also assume that the substitution effect dominates the income effect for a nonzero number of women. If the income effect dominates, the response is to stay out of the labor force.

Evidence on the effects on hours worked for those women already in the labor force is mixed, with some studies finding no effect (Eissa and Liebman 1996; Liebman 1998; Meyer and Rosenbaum 1999; and Rothstein 2007) and others finding a slight decrease in hours worked following an EITC expansion (Dickert et al. 1995; Neumark and Wascher 2001; and Saez 2010). These mixed results likely stem from evidence that EITC recipients are not well informed of the kinked structure of the EITC (Olson and Davis 1994; Smeeding et al. 2000; Ross-Phillips 2001; Romich and Weisner 2000; Maag 2005; and Chetty and Saez 2013). Supporting this, Chetty and Saez (2013) and Chetty et al. (2013) find that there is more "bunching" of incomes at kink points in the EITC schedule when recipients live in neighborhoods with higher levels of knowledge about the EITC.

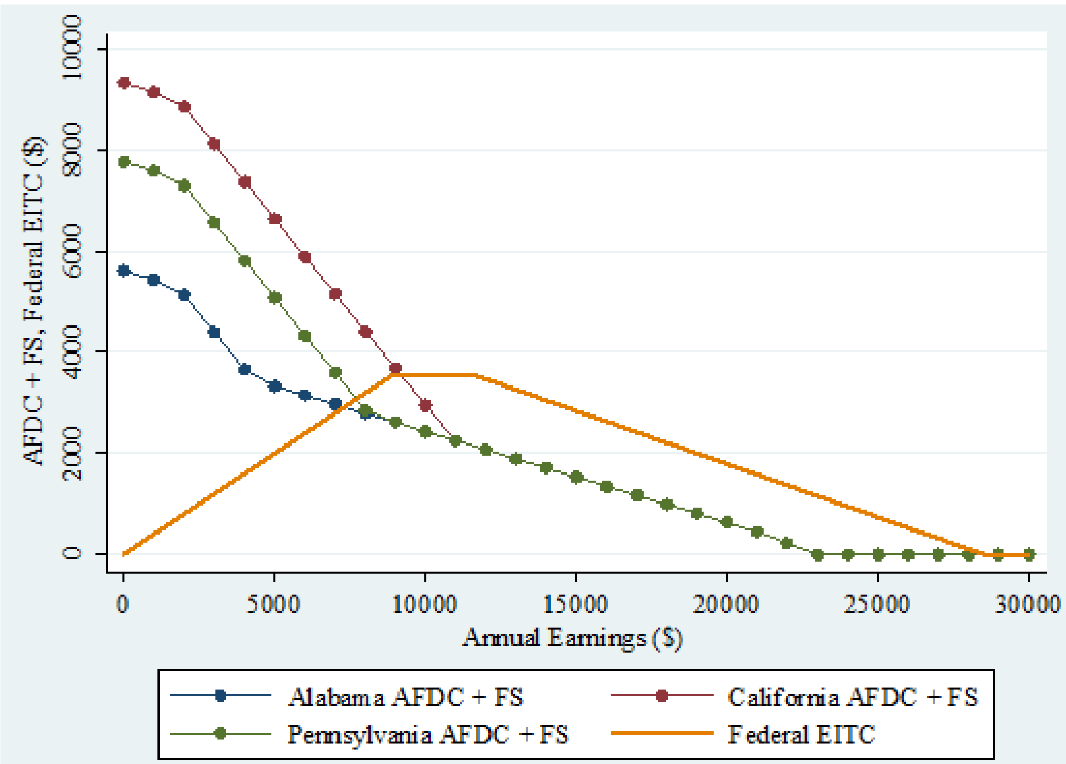
Along with changes to earnings, EITC expansions increase lump sum income through the increase in the EITC payment itself. Goodman-Bacon and McGranahan (2008) create a monthly household dataset including income, expenditures, and family structure information from the Consumer Expenditure Survey data from 1997 through 2006. They compare spending patterns of families who are eligible for the EITC to those who are not, with particular attention paid to spending in February, the modal month of EITC receipt. The authors find a small (though positive and statistically significant) effect on February expenditures for those eligible for the EITC, including relatively higher spending on food, trips, and transportation. The largest effect they find is on spending on durable goods. EITC-eligible families spend 35 percent more on new and used vehicle purchases than non-eligible families in February. EITC households also spend more on consumer electronics and household goods and appliances, though the magnitudes of the effects were small relative to the automobile purchases.

Romich and Weisner (2000) and Smeeding et al. (2000) provide qualitative evidence supporting these findings and additionally find that EITC recipients use (or plan to use) their credits to pay off bills, an expenditure category not examined in Goodman-Bacon and McGranahan (2008). Thus, it appears that EITC recipients spend EITC funds differently than typical work income, and it is possible that the effects of income increases induced by EITC changes are different from more general increases in income.

Welfare and food stamp receipt are also affected by EITC expansions for a large portion of recipients since these benefits depend on earnings. If a mother enters the labor force following an EITC expansion, the increase in earnings will increase EITC receipt but also result in lower cash welfare and food assistance benefits, where incentives vary by state. Figure 5 depicts the 1996 annual total Aid to Families with Dependent Children (AFDC) welfare cash benefits plus annual food stamps benefits for a family with a single mother and two children in Alabama, California, and Pennsylvania. I use these states to showcase the substantial variation in the generosity of AFDC programs across states. The Federal EITC schedule for 1996 is also presented in the figure (none of these states had its own EITC this year). In the phase-in region of the EITC schedule, an increase in earned income leads to an increase in EITC payments but a decrease in combined AFDC and food stamps benefits.

For a more concrete example of the potential income changes following an EITC expansion, consider a single mother with two children who is initially out of the labor force. Table 12 shows a simplified example of the various potential income sources for this family in 1996. With no positive earnings, this family's income would come entirely from monthly AFDC and food stamp payments. In Alabama, the least generous state in terms of AFDC payments, net income for this family would be about $5,616 (1996 dollars). In California, the most generous state, net income would be about $9,337, and it would be about $7,775 in Pennsylvania.

FIGURE 5
Benefit Schedule for AFDC, Food Stamps, and EITC, 1996
Single Mother with Two Children, Alabama, California, and Pennsylvania



Sources: U.S. House of Representatives, Committee on Ways and Means (1996). Joint Committee on Taxation, Ways and Means Committee (2004). Formatting partially adopted from Meyer and Rosenbaum (2000).
Notes: I assume women are past their first four months of work, have no unearned income, and claim child care expenses equal to 20 percent of earnings up to $350 per month. I assume the standard AFDC earnings disregard of $120 per month plus the child care expenses above. I assume the standard food stamps deductions of 20 percent of earnings plus $134 per month plus the child care expenses above. Shelter expenses are ignored in food stamps calculations. AFDC payments count as income in food stamps benefit calculations but not vice versa.

If instead, this mother decided to enter the labor force at a minimum wage salary – approximately $9,000 in 1996 – the family's income would change as follows in Alabama: a complete loss of the AFDC payment of $1,968, a reduction in food stamp receipt of about $1,030, an increase in Social Security taxes of about $689 (there would be no state or federal income tax liability), and a receipt of an EITC payment of about $3,566. Taken together, these changes result in an overall increase in net income of about $8,870. Similar changes occur in California and Pennsylvania with different magnitudes. Therefore, if a single mother is induced to enter the labor force following an EITC expansion, net income likely increases, but the source of income also changes significantly.

TABLE 12
Net Income Change Example, 1996
Single Mother with Two Children, Alabama, California, and Pennsylvania

State	Earnings	AFDC Payment	FS Payment	SS Tax	EITC	Net Income	Change in Net Income
Alabama	0	1,968	3,648	0	0	5,616	
	9,000	0	2,618	689	3,556	14,486	$8,870
California	0	7,284	2,053	0	0	9,337	
	9,000	1,524	2,161	689	3,556	15,553	$6,216
Pennsylvania	0	5,052	2,723	0	0	7,775	
	9,000	0	2,618	689	3,556	14,486	$6,711

Sources: U.S. House of Representatives, Committee on Ways and Means (1996). Joint Committee on Taxation, Ways and Means Committee (2004).
Notes: I assume women are past their first four months of work, have no unearned income, and claim child care expenses equal to 20 percent of earnings up to $350 per month. I assume the standard AFDC earnings disregard of $120 per month plus the child care expenses above. I assume the standard food stamps deductions of 20 percent of earnings plus $134 per month plus the child care expenses above. Shelter expenses are ignored in food stamps calculations. AFDC payments count as income in food stamps benefit calculations but not vice versa. Social Security tax is calculated at 7.65% of earnings.

For married mothers, or those mothers already in the labor force, there is likely little or no effect of an EITC expansion on labor supply. Therefore, these families should see an increase in net income that is mostly attributable to the change in their EITC payment. For example, consider a married mother with two children in 1993 and 1996, before and after OBRA 1993. Assume only the father has positive earnings of $9,000 in both years (about a minimum wage salary) and that neither parent's overall labor supply was affected by the EITC expansion. Table

13 presents the changes in the various income sources for such families in Alabama, California, and Pennsylvania.

TABLE 13
Net Income Change Example, 1993 and 1996
Married Mother with Two Children, Alabama, California, and Pennsylvania

State	Year	Earnings	AFDC Payment	FS Payment	SS Tax	EITC	Net Income	Change in Net Income
Alabama	1993	9,000	0	2,737	689	1,511	12,560	
	1996	9,000	0	3,086	689	3,556	14,954	$2,394
California	1993	9,000	1,116	2,402	689	1,511	13,341	
	1996	9,000	924	2,809	689	3,556	15,601	$2,260
Pennsylvania	1993	9,000	0	2,737	689	1,511	12,560	
	1996	9,000	0	3,086	689	3,556	14,954	$2,394

Sources: U.S. House of Representatives, Committee on Ways and Means (1996). Joint Committee on Taxation, Ways and Means Committee (2004).
Notes: I assume husbands are past their first four months of work, have no unearned income, and claim no child care expenses. I assume the standard AFDC earnings disregard of $120 per month. I assume the standard food stamps deductions of 20 percent of earnings plus $127 per month in 1993 and $134 per month in 1996 (the standard deductions). Shelter expenses are ignored in food stamps calculations. AFDC payments count as income in food stamps benefit calculations but not vice versa. Social Security tax is calculated at 7.65% of earnings.

In all three states, the change in income for these families is almost entirely driven by the increase in the EITC payment, with a small increase due to an increase in the generosity of the food stamps program during the period. In Alabama, net family income increases from $12,560 to $14,954 following the EITC expansion. $2,045 of this increase is due to the change in the EITC payment, and $349 of the increase is from the change in food stamp receipt. Therefore, after the EITC expansion, net income for these families is higher, with a large increase in lump sum income (the EITC payment).

Previous studies find that the EITC increases earnings and brings many families above the official poverty line (Neumark and Wascher 2001; Grogger 2003; Meyer 2007; Dahl et al. 2009; and Strully et al. 2010). However, earnings or poverty status do not take into account all income sources for a family, including taxes, EITC payments, and all forms of government assistance. Grogger (2003) takes into account all pre-tax income and government assistance received by families headed by a single mother. Utilizing changes in the maximum EITC credit over time, he finds that the EITC increases earnings but has no effect on income in the CPS. However, his income measure does not include taxes or EITC payments.

Meyer and Sullivan (2004) is the only study I am aware of that looks at a comprehensive measure of income that includes government assistance, EITC payments, and taxes. Using the Consumer Expenditure Survey and the Panel Study of Income Dynamics, they compare the consumption and spending patterns of single mothers from 1984 to 2000, using single women without children and married mothers as comparison groups. They find that the relative net income for single mothers is higher in the 1996 to 2000 period than in the 1984 to 1990 period, though the results are not statistically significant. The 1990s was a time period in which the EITC program was greatly expanded; however, welfare reform was also occurring during this time period, and they do not parse out the effects of this change separately. They also find that the total level of consumption for families with single mothers increased during this time period (the main focus of the study).

I expand upon the above work by examining the effect of the OBRA 1993 EITC expansion on net financial well-being separately from changes in welfare policy. I examine the effects on a comprehensive measure of net income of a family as well as the various income sources that comprise this measure, both of which have not been previously studied. I also include married families in my analysis, incorporating more families who are affected by changes in the EITC.

2.4 Data

2.4.1 Current Population Survey and Tax Liability Estimation

I use the March Current Population Survey (CPS) Annual Demographic File accessed from the Integrated Public Use Microdata Series at the Minnesota Population Center, University of Minnesota (King et al. 2010). The CPS is a repeated cross-section of between 50,000 to 62,000 households with extensive information on demographics and family structure as well as income and labor market participation from the previous calendar year. The advantages of the CPS are its large sample sizes, detailed information on income sources, including earned and unearned income, welfare, food stamp receipt, and other government assistance, and that it is nationally representative when using the provided weights. I use survey years 1991 through 2000, which correspond to effective tax years of 1990 through 1999, to analyze the effects of the 1993 OBRA expansion of the EITC.

As income measures are the key variables in this study and are self-reported in the CPS, some discussion of the validity of these measures is warranted. Only two studies of my knowledge examine the accuracy of the CPS earnings data. Bound and Krueger (1991) (BK hereafter) and Bollinger (1998) compare earnings data reported in the 1978 CPS to administrative records from the Social Security Administration. Both find that measurement error is negatively correlated with earnings, and Bollinger finds that this correlation is generated mostly by a concentration of overreporting among those with low earnings. Both also find higher measurement error in men's earnings than women's – reliability ratios for log earnings in 1997 were 0.819 and 0.924 for the two groups, respectively (BK). However, BK find that the measurement error is basically unrelated to other observable variables, suggesting that "… the mismeasurement of earnings leads to little bias when CPS earnings are on the left-hand-side of a regression."

As with self-reported earnings, transfer payment reports may also contain error. In this study, transfer payments include welfare payments from either AFDC or Temporary Assistance to Needy Families (TANF) and food stamp receipt. Food stamps are paid in-kind but treated as the cash value in my income calculations.[17] Little work has been done to directly test the accuracy of reported transfer income in the CPS, but there is some evidence of underreporting of transfer income in other data sets (Heckman and Learner 2001). However, the accuracy of transfer income is improved when looking at total transfers, as some recipients report the total fairly accurately, but do not correctly identify the specific sources of the income (Bollinger and David 1997 and Heckman and Learner 2001). For example, Bollinger and David (1997 and 2005) find that higher income households are more likely to fail to report receiving food stamps at all, but lower income households are more likely to misreport the amount of food stamps received since they confuse receipt with other welfare programs. They also find that small households and households headed by a single male are more likely to make a reporting error. Overall, it is not yet clear how errors in reported transfers affect estimates in linear regression when transfers are used as either dependent or explanatory variables (Heckman and Learner 2001).

Although the CPS contains questions about many income sources, there are not direct survey questions on taxes or EITC receipt. Using the ninth version of the National Bureau of Economic Research's TAXSIM program (Feenberg and Coutts 1993),[18] I estimate a family's tax

[17] Economic theory and prior empirical work suggest that recipients treat food stamps the same as cash income (Hoynes and Schanzenbach 2009).

[18] The program can be accessed at http://nber.org/taxsim.

liability each year, including its state and Federal EITC eligibility and payments.[19] I keep only single female heads of household or women who are married to the heads of household, where the head and spouse are both under 65. I assume the tax unit is the family, so number of children refers to biological, adopted, or step children of the mother. Therefore, income or children from unmarried partners are not taken into account in the estimation. For children with unmarried parents, I assume that the mothers claim the children. I also assume that married couples file jointly.

There are three main sources that would cause my tax estimates to have error. First is the amount of error in the self-reported income variables in the CPS. An overreporting of earnings at low levels would cause me to either over- or under-estimate the EITC payment, depending upon where the family falls in the EITC schedule that year.

Second, I implicitly assume that all qualifying households file taxes and receive the EITC and that ineligible families do not receive the credit. About 95 percent of those eligible for the EITC would either legally be required to file tax returns or would benefit from it to recover overwithheld taxes, and the IRS has a policy of notifying tax filers who appear to qualify for the EITC but do not claim it on their tax return (Scholz 1997). However, according to the Internal Revenue Service (2002), about 12.8 percent of those eligible for the EITC did not file returns for tax year 1996. About 65 percent of eligible non-filers did not make enough income to require them to file a tax return, and more than half would qualify for an EITC of less than $500.

[19] Beginning in 1992, the CPS reports values for select tax amounts including the Federal EITC payment that they also generate using the TAXSIM program. However, they do report enough tax variables to construct the total net income measure I use. For tax variables that are reported in the CPS and that I also simulate, correlations are about 0.95. I also considered using the Survey of Income and Program Participation to get around this issue since they ask questions about actual taxes, but the response rate on the tax questions for the years of interest are around 20 percent.

Therefore, the scope of misestimating EITC and taxes in my sample due to non-filers is likely small.

There is also a significant portion of EITC claims that are made by those who are not eligible for the credit. In tax year 1994, about 25 percent of all EITC claims were made by ineligible households. The largest source of the ineligible claims was due to individuals falsely claiming dependent children – about 69 percent of ineligible claims had errors of this type in 1994 (McCubbin 2000). Other ineligible claims stem from income and filing status manipulations. Noncompliance rates are highest among males filing as heads of household, suggesting that fathers falsely claim dependents when the child does not reside in the household (Liebman 2000). Therefore, in my TAXSIM calculations, I am likely understating EITC receipt since I assume no false claims; however, my underestimation won't be as severe as suggested in McCubbin (2000) since I do not include unmarried fathers in my sample.

Lastly, TAXSIM does not perfectly predict taxes and EITC payments even with reliable earnings data and perfect compliance. The CPS contains many survey questions pertinent to tax returns, but not a complete set. For example, some states allow for husbands and wives to file separately, but I assume all couples file jointly since I do not have this information (see http://nber.org/taxsim and Feenberg and Coutts (1993) for more information). However, misestimation issues due to TAXSIM are likely mitigated somewhat since I am interested in changes in taxes rather than levels and I have fairly reliable earnings data from the CPS.

2.4.2 Sample Selection and Descriptive Statistics

Following previous labor supply studies, I exclude families with mothers who are in school full-time, disabled, in the military, or whose spouses are in the military, as these women could have much different labor supply responses to tax changes than other women. To target those families who are actually affected by changes to the EITC, I include only families with children in the household and also limit the sample by income. The maximum level of real earned income allowed for EITC receipt in 1999 was about $39,500 (for a family with 2 or more children), and I limit my sample to those families with total real earned income that is less than or equal to $50,000 in a given year. Selecting the sample this way, as opposed to the typical sample restriction of less educated single mothers, allows me to include more families who are potentially eligible for the EITC.

Table 14 presents weighted summary statistics for the mothers of the full sample (with the above restrictions), the low income sample, and a sample of single mothers with a high school education or less – the most common sample restriction in EITC studies. There are 92,081 observations in the low income sample, with about 57 percent eligible for the EITC in a given year. For the single-low education sample, there are only 25,168 observations with about 60 percent eligible for the EITC. Of the 51,708 families who are EITC-eligible in the low income sample, 27,862 have married mothers, and 9,046 have a single mother with more than a high school education. The remaining 14,800 EITC-eligible families are those with a single mother with a high school education or less. Thus, limiting the sample by marital status and education misses a large portion of the population who are likely affected by changes to the EITC.[20]

[20] Limiting the sample by income is not typical in the EITC literature, as changes in the EITC affect labor supply decisions and, thus, income. In a repeated cross section, limiting the sample by income could cause composition changes in the sample over time if families with pre-tax income above the income cutoff reduce their earnings following an expansion in order to receive the credit. However, for the sample to change composition in this way, families with real earnings above $50,000 would have to reduce their earnings (or their reported earnings) by more than $10,000 in response to the 1993 EITC expansion in order to receive even a small credit. As the maximum Federal credit was less than $5,000 in 1999, I find this unlikely to occur.

TABLE 14
Summary Statistics – Various Subsamples, 1990 - 1999

VARIABLES	All	Low Income <= $50K	Single, <= HS Education
Black	0.13	0.20	0.35
	(0.34)	(0.40)	(0.48)
White	0.82	0.75	0.62
	(0.38)	(0.43)	(0.49)
Asian	0.04	0.03	0.01
	(0.18)	(0.17)	(0.12)
Age	37.15	35.37	34.97
	(8.26)	(8.94)	(9.30)
Married	0.77	0.55	0.00
	(0.42)	(0.50)	(0.00)
Highest Grade Completed	13.00	11.95	11.07
	(2.60)	(2.52)	(1.83)
Number Children	1.96	2.00	1.96
	(0.99)	(1.09)	(1.12)
EITC Eligible	0.26	0.57	0.60
	(0.44)	(0.50)	(0.49)
In Poverty	0.14	0.31	0.52
	(0.35)	(0.46)	(0.50)
In Labor Force	0.73	0.65	0.69
	(0.45)	(0.48)	(0.46)
Family Earned Income	64.60	23.23	14.36
	(59.63)	(15.94)	(18.91)
EITC Payment	0.48	1.03	1.17
	(1.05)	(1.34)	(1.42)
Welfare + Food Stamps	0.78	1.67	3.37
	(2.78)	(3.89)	(5.02)
Family Net Income	54.02	28.61	22.40
	(39.43)	(17.50)	(17.28)
Observations	193,613	92,081	25,168

Notes: Summary statistics of various subsamples of mothers and their families in the March CPS Annual Demographic File. Standard deviations in parenthesis. Means and standard deviations weighted by CPS household weights. See text for sample selection and description.

For this reason, I prefer the low income cutoff and will use that sample in my remaining analyses. However, I also present the main results for the single-low education sample in the appendix for comparison to previous findings. About 20 percent of mothers in the low income sample are black and 55 percent are married. The average age of mothers in the sample is around 35 years old, and the highest grade completed is near a high school education. There are about 2 children in each family.

To answer the central question in this paper of how the EITC affects the financial well-being of its recipients, I examine six key outcome variables. I first examine the official poverty status of the family reported in the CPS. This measure is based on need and all pre-tax cash income from all sources including government assistance of all family members in the household; however, it does not include noncash government assistance (such as, food stamps, housing subsidies) or EITC payment and taxes. About 31 percent of the low income sample is in poverty.

I next consider the effect of the EITC expansion on the labor force participation of mothers for comparison with previous studies. As described above, changes in the EITC incentives are likely one of the driving forces in income changes following EITC expansions. To measure these income changes driven through labor supply choices, I also include the earned income of the family, which includes wages, salary, and income from an owned business or farm of the head of household and spouse (if married). About 65 percent of mothers in the low income sample are in the labor force, and average annual earned income is about $23,200.

I also examine welfare and EITC payments separately from other income sources to determine the extent to which EITC expansions affect the substitution of family income from welfare to earnings and EITC payments. Welfare payments include yearly AFDC or TANF cash welfare payments plus the dollar value of yearly food stamp receipt as reported in the CPS.[21] Families receive about $1,700 in combined welfare and food stamps in the low income sample. The EITC payment is estimated with TAXSIM as described above, and the average credit is about $1,000.

Lastly, I construct a comprehensive net family income measure to examine changes to overall family financial resources. This measure includes the total family income variable from the CPS that includes total pre-tax cash income from all sources including government assistance[22] to which I add the cash value of food stamps and deduct taxes estimated with TAXSIM, including EITC payments. Average net family income is about $28,600 in my sample.

2.5 Methodology

I exploit exogenous variation in EITC generosity over time and across family size resulting from the OBRA 1993 tax reform to estimate causal responses to the EITC expansion. OBRA 1993 increased real EITC payments and differentiated the credit by number of children for the first time. After this expansion, families with two or more children are eligible for a much larger credit than families with only one child – a relative increase in the real maximum EITC of about $1,900 (see Figure 4 for illustration). As is popular in the EITC literature, I employ a DiD approach comparing families with two or more children to those with only one child before and after the 1993 expansion.

[21] The annual food stamp value is the only income variable defined at the household rather than family level in the CPS because the food stamp unit is the household rather than the family.
[22] Cash income sources comprising the total family income variable include the following: income from salary and wages, income from a farm or business, Social Security payments, AFDC/TANF payments, pension or retirement income, Supplemental Security Income, interest income, unemployment payments, workers' compensation, Veterans' Administration payments, survivors' benefits, disability income, rental income, educational assistance, alimony and child support payments, regular cash assistance from friend or relatives outside the household, and a residual category for any other pre-tax income.

I estimate the following model:

$$y_{ist} = \alpha + After * 2plus_{ist}\beta_1 + 2PlusChildren_{ist}\beta_2 + \boldsymbol{Welfare_{st}\beta_3} + \boldsymbol{X_{ist}\beta_4} + \delta_t + \gamma_s + \varepsilon_{ist} \quad (1)$$

where i indexes the mother (family), s indexes state, t indexes year, and ε_{ist} is an idiosyncratic error term. y_{ist}, the outcome of interest, is a measure of financial well-being described above – an indicator for whether the mother is in poverty, an indicator for whether a family is in the labor force, real family earned income, real EITC payment, real combined welfare and food stamp receipt, or real net family income. $\boldsymbol{X_{ist}}$ is a row vector of controls including indicators for mother's marital status, race, and highest grade completed, mother's age and its square, and an indicator for whether there are children under age 5 in the household. $2PlusChildren_{ist}$is an indicator for whether there are two or more children in the family. I also include state and year fixed effects, γ_s and δ_t. $After * 2plus_{ist}$ is a dummy variable indicating that the year is 1996 or later (post OBRA 1993) and that there are two or more children in the family. The before period in the analyses is 1990 through 1993, with 1994 and 1995 being omitted, as the tax change was not yet fully phased in. I cluster standard errors at the state level and use CPS household weights in all regressions and summary statistics.

I am effectively comparing families with two or more children (the treatment group) to those with one child (the control group) in 1990-1993 (before period) and 1996-1999 (after period). The underlying identifying assumption is that there is nothing unobservable that differentially affects the treatment and control groups other than the policy change in this time frame. Therefore, β_1, the parameter of interest, captures the relative increase in the given outcome variable for the treatment to control group that is attributable to the expansion in the

EITC. Note that time-invariant differences in the outcome variable for the two groups do not threaten the validity of this strategy.

Although I am able to control for a rich set of controls, there still could be other unobservable factors or contemporaneous policy changes that differentially affect the two groups. One possibility is changes to other welfare programs. Between 1993 and 1996, 43 states received waivers to experiment with changes to the welfare program, AFDC. These waivers generally required work, set time limits for assistance, or increased work incentives (Meyer and Rosenbaum 2000). In 1996, TANF replaced AFDC, and, like the EITC, increased the emphasis on work as well as gave states greater discretion in designing their programs (Rowe 2000). To address these changes in welfare policy over the period, I include the row vector $\boldsymbol{Welfare_{st}}$, which contains the maximum welfare benefit in state s in year t for a family of three as well as an indicator for whether any time limits or work requirements for welfare receipt had been put in place.[23]

Along with the basic DiD results, I test for differential time trends in the outcome variables for the treatment and control groups using an event study design by estimating the following equation:

$$y_{ist} = \alpha + \sum_{y=1990}^{1992} 1[t = y] * 2plus_{ist}\theta_y + \sum_{y=1994}^{1999} 1[t = y] * 2plus_{ist}\pi_y$$

$$+ 2PlusChildren_{ist}\tau_1 + \boldsymbol{Welfare_{st}\tau_2} + \boldsymbol{X_{ist}\tau_3} + \delta_t + \gamma_s + \varepsilon_{ist} \qquad (2)$$

[23] I obtained the welfare variables from the Urban Institute's Welfare Rules Database (http://anfdata.urban.org/wrd/WRDWelcome.cfm) and from data used in Meyer and Rosenbaum (2001) that Bruce D. Meyer generously provided.

where $1[t = y]$ is an indicator function that the year is y, and all other variable definitions are the same as above. Thus, the coefficient estimates on the set of θ_y represent the pre-OBRA 1993 trend in the outcome variable conditional on observables for families with two or more children relative to one-child families, with 1993 being the omitted year. Similarly, the coefficient estimates on the set of π_y represent the relative trend after OBRA 1993 went into effect. If the pre-trend is not flat, the above DiD estimates may not be valid. For example, say the relative maternal labor market participation in the treatment group compared to the control group was trending upward before the tax change. If a positive effect is found on the treatment variable in equation (1), it is not clear whether this is a result of the increase in relative EITC benefits or just a continuation of the pre-trend in the outcome variable.

2.6 Results

2.6.1 Difference-in-Differences Results

Table 15 presents summary statistics by treatment group before and after OBRA 1993 as well as the unconditional DiD estimates for the six outcome variables for the low income sample. I find strong evidence that the EITC substantially improves the financial well-being of its recipients. OBRA 1993 significantly reduces the relative likelihood of being in poverty of the treatment group by about 2.1 percentage points – a 6.8 percent increase from the sample poverty rate of 31 percent. Although not statistically significant, I find a positive effect of OBRA 1993 on labor force participation. This result is consistent with previous work, as I am looking at single and married women together – that is, the previous finding of a large significant increase in labor force participation of single mothers will be somewhat masked by the inclusion of married women in the analysis (I examine the two groups separately below). I find highly statistically significant results for the remaining outcomes. The EITC expansion increases earnings by about $870 (3.7 percent of sample mean), increases EITC payments by about $739 (71.1 percent), decreases combined welfare and food stamp receipt by about $551 (33.0 percent), and increases total net family income by about $950 (3.3 percent).

TABLE 15
Summary Statistics by Treatment and Unconditional Difference-in-Differences Estimates
Families w/ Real Earned Income <= $50,000

	One Child		Two Plus Children		
VARIABLES	1990-93	1995-99	1990-93	1995-99	Diff-in-Diff
In Poverty	0.242	0.223	0.382	0.342	-0.021***
	(0.004)	(0.004)	(0.003)	(0.004)	(0.008)
In Labor Force	0.662	0.742	0.569	0.658	0.009
	(0.004)	(0.005)	(0.004)	(0.004)	(0.008)
Family Earned Income	22.840	24.170	21.990	24.190	0.870***
	(0.145)	(0.159)	(0.118)	(0.128)	(0.277)
EITC Payment	0.579	1.042	0.567	1.769	0.739***
	(0.007)	(0.012)	(0.005)	(0.015)	(0.021)
Welfare + Food Stamps	1.077	0.630	2.580	1.582	-0.551***
	(0.025)	(0.021)	(0.035)	(0.030)	(0.057)
Family Net Income	27.310	28.150	28.340	30.130	0.950***
	(0.154)	(0.187)	(0.118)	(0.143)	(0.305)
Observations	16,470	11,820	26,712	19,574	

Notes: Weighted by CPS household weights. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 16 presents the main regression-adjusted DiD estimates from equation (1). I find treatment effects fairly similar in size to the unconditional estimates. OBRA 1993 reduces the relative likelihood of poverty by about 1.3 percentage points (4.2 percent of sample mean) for families with two or more children and increases labor force participation by about 1.7

percentage points (2.6 percent). The expansion increases relative earnings by about $471 (2.0 percent), increases EITC payments by about $742 (72.0 percent), decreases welfare and food

TABLE 16
Main Difference-in-Differences Results
Families w/ Real Earned Income <= $50,000

VARIABLES	In Poverty	In Labor Force	Earned Income	EITC Payment	Welfare + FS	Net Income
*After * Two Plus*	-0.013*	0.017**	0.471***	0.742***	-0.464***	0.527**
	(0.007)	(0.006)	(0.173)	(0.020)	(0.052)	(0.217)
Two Plus Children	0.131***	-0.065***	-2.159***	0.001	1.388***	2.263***
	(0.004)	(0.005)	(0.167)	(0.015)	(0.094)	(0.170)
Welfare Max Benefit	-0.006*	-0.010***	0.091	-0.025	0.064	0.672***
	(0.003)	(0.003)	(0.115)	(0.015)	(0.053)	(0.159)
Time Limits on Welfare	-0.000	0.001	-0.274	0.023	0.026	-0.115
	(0.012)	(0.010)	(0.406)	(0.035)	(0.111)	(0.462)
Married	-0.241***	-0.146***	13.81***	-0.330***	-1.919***	8.330***
	(0.005)	(0.008)	(0.217)	(0.056)	(0.179)	(0.176)
Black	0.046**	0.068***	-0.148	0.039	0.307	-1.648***
	(0.022)	(0.025)	(0.680)	(0.050)	(0.197)	(0.550)
White	-0.073***	0.058**	2.841***	-0.001	-0.755***	1.827***
	(0.021)	(0.023)	(0.646)	(0.042)	(0.190)	(0.478)
Asian	0.024	0.027	-1.741*	0.053	0.005	-0.734
	(0.024)	(0.023)	(1.012)	(0.060)	(0.391)	(0.768)
Children Under Five	0.087***	-0.156***	-1.677***	-0.067***	0.797***	-2.318***
	(0.005)	(0.005)	(0.177)	(0.019)	(0.100)	(0.179)
Age	-0.020***	0.022***	1.451***	-0.021***	-0.077***	-0.096
	(0.002)	(0.002)	(0.079)	(0.004)	(0.016)	(0.079)
Age^2	0.000***	-0.000***	-0.019***	0.000**	0.001***	0.008***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
Year Fixed Effects	x	x	x	x	x	x
State Fixed Effects	x	x	x	x	x	x
Highest Grade FEs	x	x	x	x	x	x
Observations	74,572	74,572	74,572	74,572	74,572	74,572
R-squared	0.205	0.129	0.287	0.188	0.212	0.196

Notes: Robust standard errors (clustered at the state level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Each column represents a separate WLS regression using CPS household weights for the various outcomes.

stamp receipt by about $464 (27.8 percent), and increases total net family income by about $527 (1.8 percent). All estimates are statistically significant at the ten percent level or better. These results are consistent with the predictions laid out in Section 2.3: Overall, the EITC expansion results in a positive increase in labor force participation and earnings. As a result, welfare and food stamp receipt declines and EITC receipt increases, with a combined positive effect on net income of the family.

Table 17 presents the coefficient estimate for the treatment variable, $After * 2plus_{ist}$, for various subgroups of the sample. Each line represents a different set of estimates of β_1 obtained from running a separate regression for each outcome on the given subgroup. The first row of the table reproduces the results on the full low income sample from Table 16, and the second panel presents the results separately by marital status. Consistent with previous studies (Dickert et al. 1995; Eissa and Liebman 1996; Ellwood 2000; Meyer and Rosenbaum 2000 and 2001; Neumark and Wascher 2001; Grogger 2003; Eissa and Hoynes 2004; Hotz et al. 2006; Rothstein 2007; and Adireksombat 2010), I find OBRA 1993 had a large, statistically significant positive effect for single mothers (about 4.7 percentage points) and cannot reject a zero effect for married mothers.

For comparison to previous studies, I also present the results for single mothers with a high school education or less in Table 22 in the appendix. For this group, I find that OBRA 1993 increased the relative labor force participation for mothers with two or more children by about 5.1 percentage points, statistically significant at the one percent level. Though not shown, the estimate for married mothers with a high school education or less is -1.2 percentage points with a standard error of .9. These results are very comparable with previous labor supply studies (Dickert et al. 1995; Eissa and Liebman 1996; Meyer and Rosenbaum 2000 and 2001; Neumark and Wascher 2001; Grogger 2003; Eissa and Hoynes 2004; Hotz et al. 2006; and Rothstein 2007), especially those examining OBRA 1993 using the CPS (Adireksomdat 2010 and Ellwood 2000). For example, Adireksomdat (2010) finds that the relative labor force participation of single mothers with a high school education or less with two or more children compared to those with one child increased by 5.2 percentage points as a result of OBRA 1993 (compare to my estimate of 5.1 percentage points).

TABLE 17
Difference-in-Differences Results by Subgroups
Families w/ Real Earned Income <= $50,000

VARIABLES	In Poverty	In Labor Force	Earned Income	EITC Payment	Welfare + FS	Net Income
All	-0.013*	0.017**	0.471***	0.742***	-0.464***	0.527**
	(0.007)	(0.006)	(0.173)	(0.020)	(0.052)	(0.217)
Single	-0.016	0.047***	0.894***	0.868***	-0.972***	0.472
	(0.011)	(0.011)	(0.307)	(0.043)	(0.085)	(0.365)
Married	-0.018**	0.015	0.391	0.718***	-0.350***	0.461
	(0.008)	(0.012)	(0.320)	(0.043)	(0.054)	(0.440)
White	-0.006	0.007	0.019	0.713***	-0.292***	0.381
	(0.007)	(0.007)	(0.173)	(0.030)	(0.051)	(0.265)
Minority	-0.030**	0.042***	1.880***	0.806***	-1.040***	0.857
	(0.015)	(0.013)	(0.485)	(0.052)	(0.118)	(0.563)

Notes: Robust standard errors (clustered at the state level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Each cell represents the coefficient estimate and standard error for the *After * TwoPlus* variable for a separate WLS regression using CPS household weights for the various outcomes and subgroups with full set of controls from Table 16.

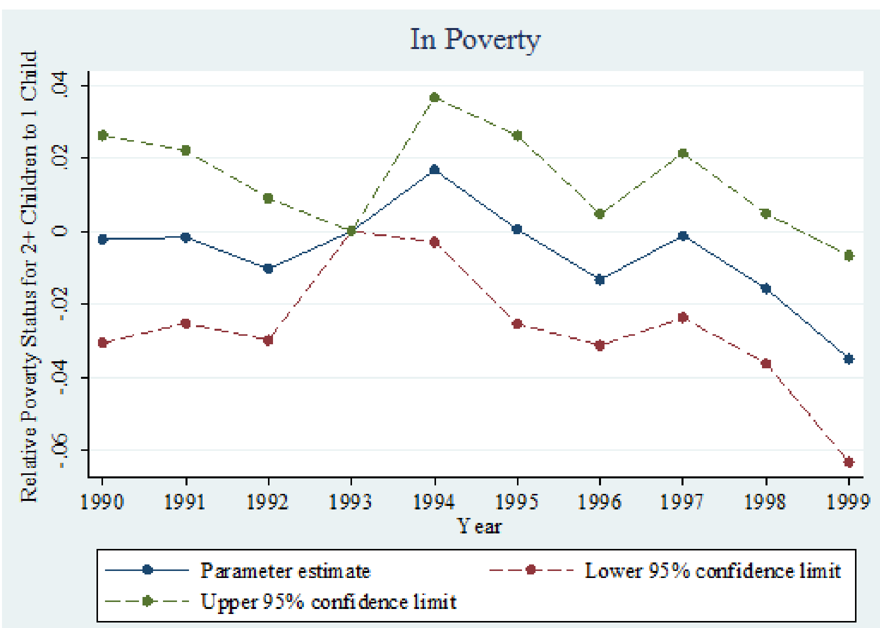
Single mothers also have statistically different estimates from those for married mothers (one percent level) for EITC payment and combined welfare and food stamp receipt. Though not statistically different from one another, I find a higher point estimate for single mothers on earned income but a lower point estimate on poverty reduction compared to married mothers. Consistent with the above predictions, single women experience a large increase in labor supply, earnings, and EITC receipt following an EITC expansion, but see a large loss in welfare and food stamp receipt. The main effect for married women, who initially have much higher earnings, is driven through the higher EITC payment, as the overall labor supply for this group is not affected much following an expansion. I do estimate a reduction in welfare and food stamp receipt for married women as well, but the point estimate is much smaller than that for single women as would be expected. Though the sources of income changes vary, the overall effect on net family income does not appear to differ much for these two groups. Therefore, families with married mothers should also be considered in EITC studies along with those with low-educated single mothers that are typically considered since their financial resources are significantly affected as well.

The last panel in Table 17 presents the coefficient estimate for $After * 2plus_{ist}$ separately for white and non-white families to further explore changes by income level, where I am considering minority status a proxy for income. In the sample, white families have average earnings of $23,974 with about 13 percent having zero earnings prior to the OBRA 1993 expansion. Non-white families have average earnings of $16,992 with about 27 percent having zero earnings before the expansion. I find that OBRA 1993 had a significantly larger effect on improving the financial well-being of minority families. All point estimates are statistically different from one another for white versus minority families at the ten percent significance level or better with the exception of the net income variable. As minority families have much lower initial earnings and a large proportion at zero earnings, they have a much larger labor supply response. This results in larger increases to earnings, EITC receipt, and overall net income for this group.

2.6.2 Event Study Results

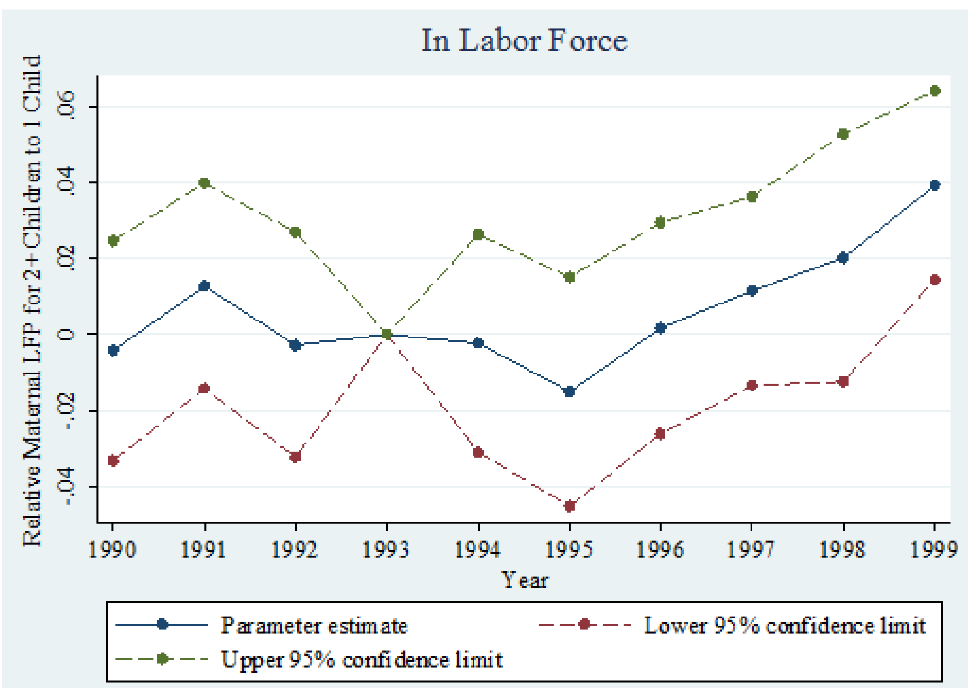
Figures 6 through 11 present graphs of the coefficient estimates and their 95 percent confidence intervals for the set of θ_y and π_y from the event study design in equation (2) for the various outcomes in the low income sample. For example, in Figure 6, these estimates represent the regression-adjusted relative poverty rates for mothers with 2 or more children compared to those with one child each year, with 1993 being the omitted year. To add validity to the above DiD estimates, the event study results should show a flat trend in the relative outcome variables of the treatment and control groups before 1993.

FIGURE 6
Event Study Results – In Poverty
Families w/ Real Earned Income <= $50,000



In Figure 6 we see that, prior to 1993, the relative trend in the poverty rate of the treatment and control groups is basically flat and none of the coefficient estimates are statistically different from zero. In 1996, once the tax change is fully phased in, there is a gradual decline in the coefficient estimates, with the 1999 estimate being statistically significant. In 1999, the relative poverty rate was about 3.5 percentage points lower for the treatment group compared to the control than in 1993. To compare to the DiD results above, we would compare the average coefficient estimate from 1996 through 1999 to the average in 1990 through 1993 in the event study – a relative decline in poverty of about 1.26 percent in the DiD and 1.25 percent in the event study. The estimates are almost identical as expected since these are very similar identification strategies.

FIGURE 7
Event Study Results – In Labor Force
Families w/ Real Earned Income <= $50,000



The remaining graphs in Figures 7 through 11 present the event study results for the other five outcome variables. As with the poverty results, the other outcomes also have a fairly flat pre-trend in the relative outcomes followed by a gradual increase (decrease in the case of welfare and food stamp receipt) once OBRA 1993 is phased in. As with poverty status, the magnitudes of the estimates for the other outcomes are almost identical to the DiD estimates when averaging across the periods before and after 1993.

FIGURE 8
Event Study Results – Family Earned Income
Families w/ Real Earned Income <= $50,000

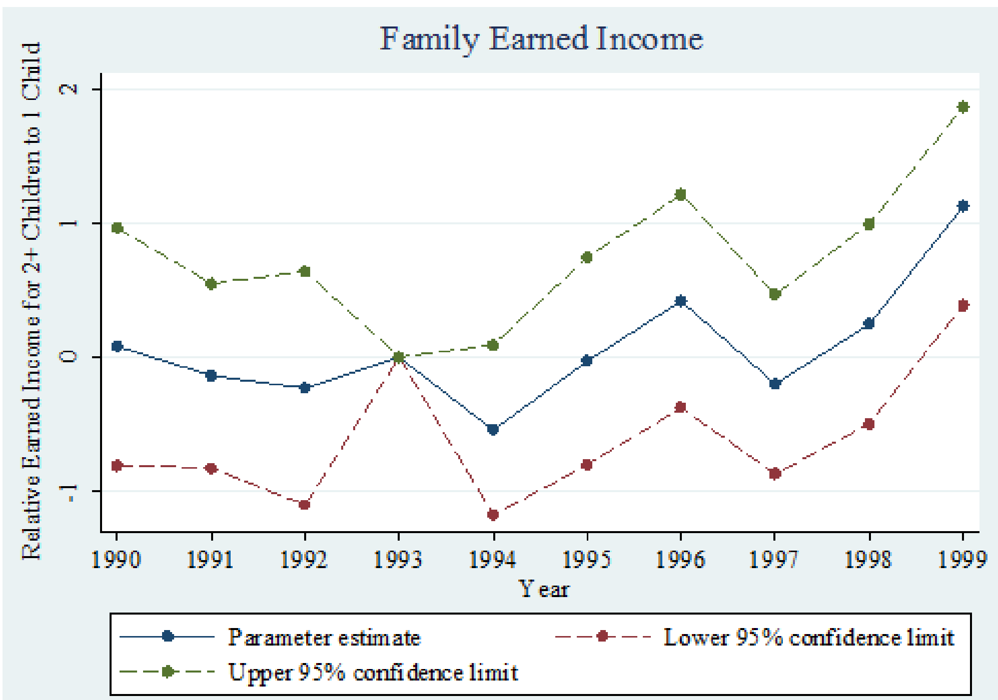
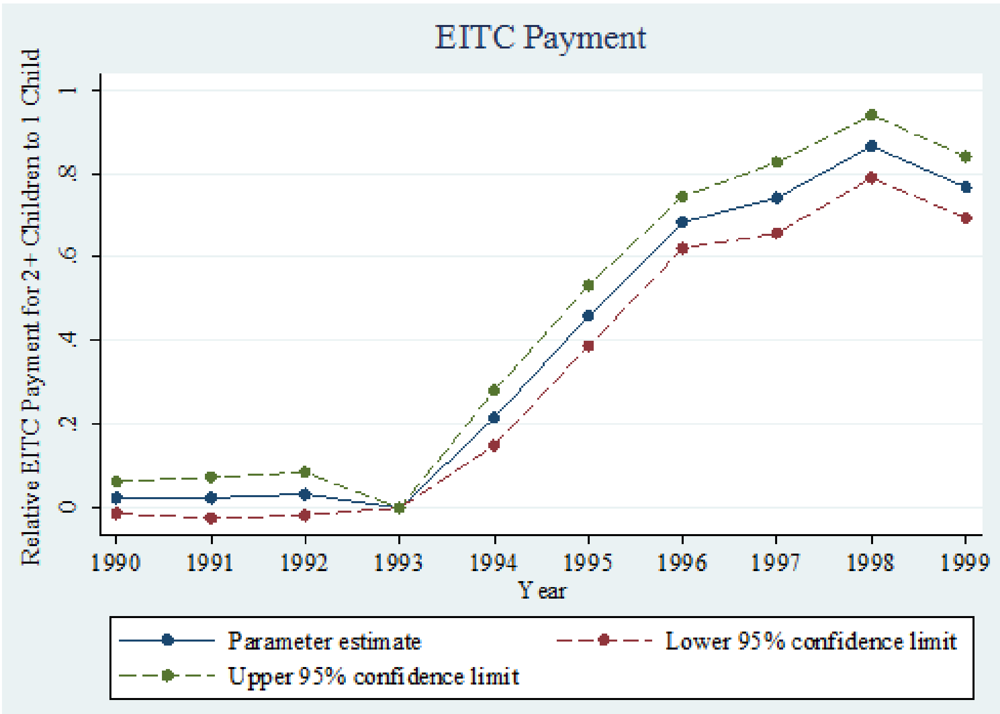
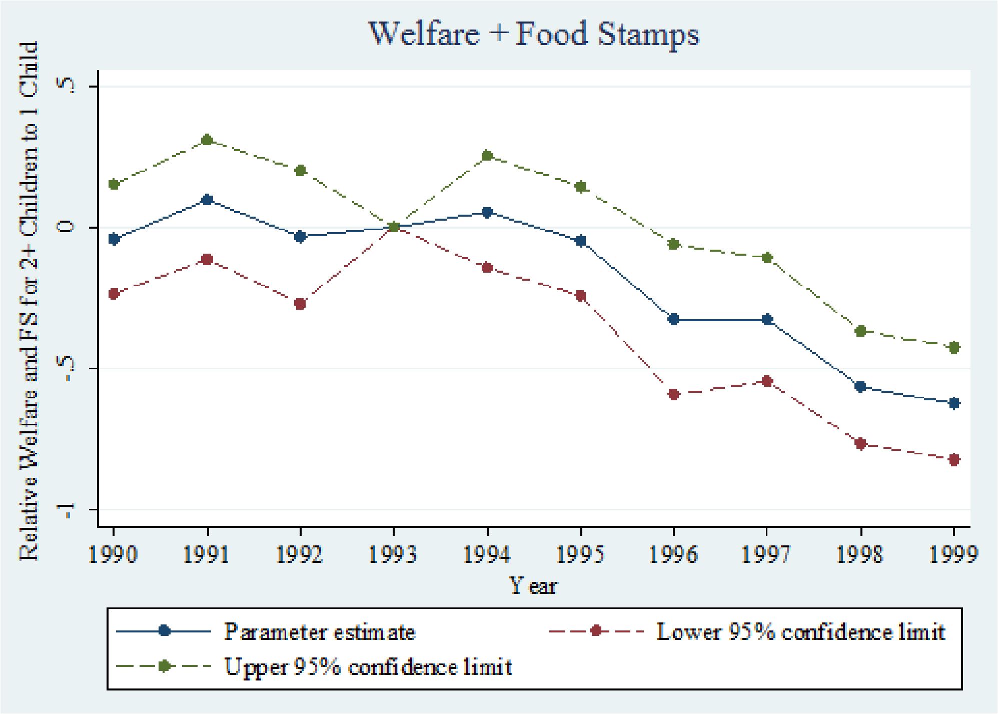


FIGURE 9
Event Study Results – EITC Payment
Families w/ Real Earned Income <= $50,000



It appears that families were not reacting to OBRA 1993 immediately. Labor supply responses gradually occur after 1996 once the change was fully phased in, and relative EITC payments also continue to increase for families with 2 or more children compared to those with one child after 1996. This suggests that families might not be fully aware of EITC expansions immediately. Relative net family income doesn't increase until 1998 due to this delayed response. As a result, the DiD results above might actually understate the effect of OBRA 1993 on net income since the after period (1996 to 1999) includes years when responses weren't yet fully realized.

FIGURE 10
Event Study Results – Welfare and Food Stamps
Families w/ Real Earned Income <= $50,000

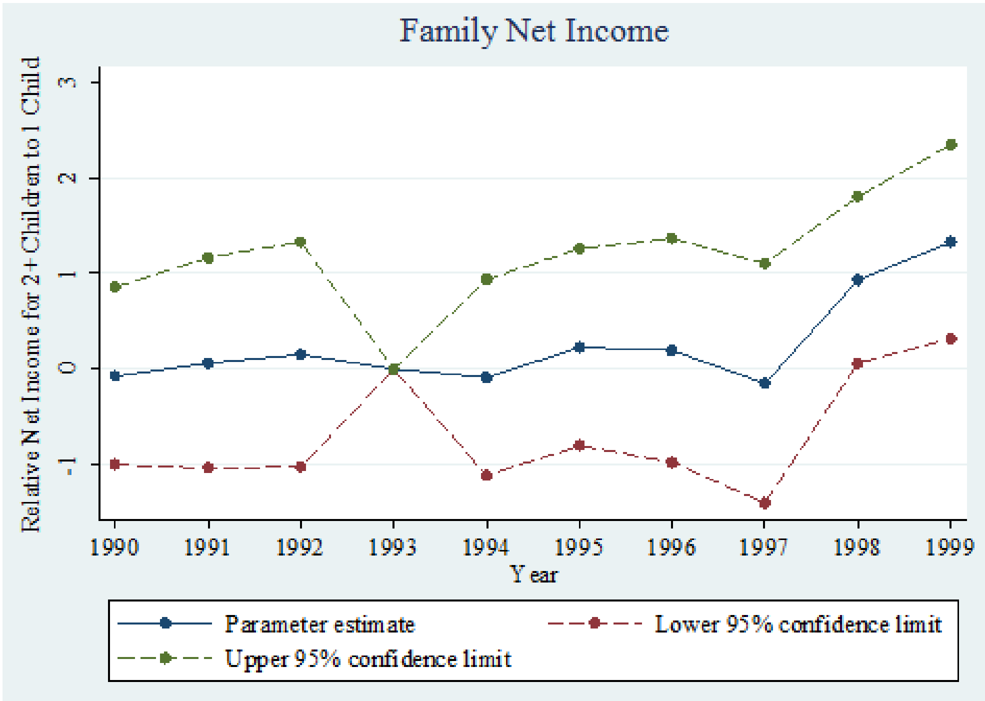


2.7 Summary and Conclusions

Using the CPS, I examine the effects of the OBRA 1993 expansion of the EITC on the financial well-being of its recipients. I use a low income sample that includes all families potentially affected by EITC changes as opposed to the typical approach of limiting the sample to low educated single mothers. This study is the first of my knowledge to examine the effects of EITC expansions on a comprehensive measure of total net family income as well as the various income sources that comprise this measure.

Using a popular DiD framework comparing families with two or more children to those with one child before and after OBRA 1993, I find that this EITC expansion significantly improves the financial situation of its recipients. Following a relative increase in the maximum

FIGURE 11
Event Study Results – Family Net Income
Families w/ Real Earned Income <= $50,000



EITC of about $1,900 (2008 dollars), treated families increase relative labor force participation by about 1.7 percentage points and are 1.3 percentage points less likely to be in poverty. OBRA 1993 also increases relative earnings by about $471, increases EITC payments by about $742, and increases total net family income by about $527 for families with two or more children. As only 63 percent of families in my sample are eligible for the EITC after the expansion, the increases in net income experienced by those actually receiving the EITC are much larger than the point estimates suggest. I also find larger effects for single mothers and minority families, suggesting that the program is well-targeted at the most disadvantaged families.

Overall, the EITC appears to be an effective policy for encouraging work and improving the financial status of low income families. Not only does net income increase after an expansion, but the sources of income also change, shifting income from welfare and food stamp receipt to earnings and EITC payments. Social stigma costs of welfare receipt (Levine and Zimmerman 2005) and the lump sum nature of EITC payments (Romich and Weisner 2000; Smeeding et al. 2000; and Goodman-Bacon and McGranahan 2008) might mean that families benefit more from the overall income increase following an EITC expansion than from a typical increase in income. The findings in this paper add support to previous work finding large effects of the EITC on health and educational outcomes (Baker 2008; Strully et al. 2010; Chetty et al. 2011a; Dahl and Lochner 2012; Hoynes et al. 2012; Maxfield 2013; Michelmore 2013; Evans and Garthwaite 2014; and Manoli and Turner 2014).

CHAPTER 3

An Evaluation of Empirical Bayes' Estimation of Value-Added Teacher Performance Measures

Coauthored with Cassandra M. Guarino, Mark D. Reckase, Paul Thompson, and Jeffrey M. Wooldridge

3.1 Introduction

Empirical Bayes' (EB) estimation of teacher effects has gained recent popularity in the value-added research community (see, for example, McCaffrey et al. 2004; Kane and Staiger 2008; Chetty, Friedman, and Rockoff 2011b; Corcoran, Jennings, and Beveridge 2011; and Jacob and Lefgren 2005, 2008). Researchers motivate the use of EB estimation as a way to decrease classification error of teachers, especially when limited data is available to compute value-added estimates. When there are only a small number of students per teacher, teacher value-added estimates can be very noisy. EB estimates of teacher value-added reduce the variability of the estimates by shrinking them toward the average estimated teacher effect in the sample and, therefore, are often referred to as "shrinkage estimators." As the degree of shrinkage depends on class size, estimates for teachers with smaller class sizes are more affected, potentially helping with the misclassification of these teachers. In addition, EB estimation may be less computationally demanding than methods that view the teacher effects as fixed parameters to estimate.

Despite the potential shrinkage benefits of EB estimation, the estimated teacher effects can suffer from severe bias under nonrandom teacher assignment. By treating the teacher effects as random, EB estimation assumes that teacher assignment is uncorrelated with factors that predict student achievement -- including observed factors such as past test scores. While the bias (technically, the inconsistency) disappears as the number of students per teacher increases -- because the EB estimates converges to the so-called fixed effects estimates -- the bias still can be important for the kinds of data used to estimate teacher VAMs. This is because the EB estimators of the coefficients on the covariates in the model are inconsistent for fixed class sizes as the number of classrooms grows. By contrast, estimators that include the teacher assignment indicators along with the covariates in a multiple regression analysis are consistent (as the number of classrooms grows) for the coefficients on the covariates. This generally leads to less bias in the estimated teacher VAMs under nonrandom assignment without many students per teacher.

In this paper we address the following research questions: (1) How does the performance of EB estimators compare with that of estimators that treat the teacher effect as fixed under random teacher assignment and various nonrandom assignment scenarios? (2) Are there cases where it is beneficial to use an EB-type approach to shrink estimates of teacher fixed effects? (3) How do recently proposed simplified versions of EB perform?

3.2 Empirical Bayes' Estimation

There are several ways to derive empirical Bayes' estimators of teacher value added. We adopt a so-called "mixed estimation" (ME) approach, as in Ballou, Sanders, and Wright (2004), because it is fairly straightforward and does not require delving into Bayesian estimation methods. Our focus is on estimating teacher effects grade by grade. Therefore, we assume either that we have a single cross section or multiple cohorts of students for each teacher. We do not include cohort effects; multiple cohorts are allowed by pooling students across cohorts for each teacher.

Let y_i denote a measure of achievement for student i, randomly drawn from the population. This measure could be a test score or a gain score. Suppose there are G teachers and the teacher effects are b_g, $g = 1, \ldots, G$. In the mixed effects setting, these are treated as random variables as opposed to fixed population parameters. Viewing the b_g as random variables independent of other observable factors affecting test scores has consequences for the properties of EB estimators.

Typically VAMs are estimated controlling for other factors, which we denote by a row vector $\boldsymbol{x_i}$. These factors include student demographics and, in some cases, prior test scores. We assume the coefficients on these covariates are fixed parameters. We can write a mixed effects linear model as

$$y_i = \mathbf{x}_i\gamma + \mathbf{z}_i\mathbf{b} + u_i, \tag{1}$$

where $\boldsymbol{z_i}$ is a $1 \times G$ row vector of teacher assignment dummies, $\boldsymbol{b}$ is the $G \times 1$ vector of teacher effects, and u_i contains the unobserved student-specific effects. Because a student is assigned to one and only one teacher, $z_{i1} + z_{i2} + \cdots + z_{iG} = 1$. Equation (1) is an example of a "mixed model" because it includes the usual fixed population parameters γ and the random coefficients $\boldsymbol{b}$. Even if there are no covariates, $\boldsymbol{x_i}$ typically includes an intercept. If $\boldsymbol{x_i\gamma}$ is only a constant, so $\boldsymbol{x_i}\gamma = \gamma$, then γ is the average teacher effect and we can then take $E(\boldsymbol{b}) = \boldsymbol{0}$. This means that b_g is the effect of teacher g net of the overall mean teacher effect.

Equation (1) is written for a particular student i so that teacher assignment is determined by the vector $\boldsymbol{z_i}$. A standard assumption is that, conditional on $\boldsymbol{b}$, (1) represents a linear conditional mean:

$$E(y_i|\mathbf{x}_i, \mathbf{z}_i, \mathbf{b}) = \mathbf{x}_i\gamma + \mathbf{z}_i\mathbf{b} \tag{2}$$

which follows from

$$E(u_i|\mathbf{x}_i, \mathbf{z}_i, \mathbf{b}) = 0. \tag{3}$$

An important implication of (3) is that u_i is uncorrelated with $\boldsymbol{z_i}$, so that teacher assignment is not systematically related to unobserved student characteristics once we have controlled for the observed factors in $\boldsymbol{x_i}$.

If we assume a sample of N students assigned to one of G teachers we can write (1) in matrix notation as

$$\mathbf{y} = \mathbf{X}\gamma + \mathbf{Z}\mathbf{b} + \mathbf{u}, \tag{4}$$

where $\boldsymbol{y}$ and $\boldsymbol{u}$ are $N \times 1$, $\boldsymbol{X}$ is $N \times K$, and $\boldsymbol{Z}$ is $N \times G$. In order to obtain the best linear unbiased estimator (BLUE) of γ and the best linear unbiased predictor (BLUP) of $\boldsymbol{b}$, we assume that the covariates and teacher assignments satisfy a strict exogeneity assumption:

$$E(u_i|\mathbf{X}, \mathbf{Z}, \mathbf{b}) = 0,\ i = 1, ..., N. \tag{5}$$

An implication of assumption (5) is that inputs and teacher assignment of *other* students does not affect the outcome of student i.

Given assumption (5) we can write the conditional expectation of $\boldsymbol{y}$ as

$$E(\mathbf{y}|\mathbf{X}, \mathbf{Z}, \mathbf{b}) = \mathbf{X}\gamma + \mathbf{Z}\mathbf{b} \tag{6}$$

In the EB literature a standard assumption is

$$\mathbf{b} \text{ is independent of } (\mathbf{X}, \mathbf{Z}), \tag{7}$$

in which case

$$E(\mathbf{y}|\mathbf{X}, \mathbf{Z}) = \mathbf{X}\gamma + \mathbf{Z}E(\mathbf{b}|\mathbf{X}, \mathbf{Z}) = \mathbf{X}\gamma = E(\mathbf{y}|\mathbf{X}) \tag{8}$$

because $E(\boldsymbol{b}|\boldsymbol{X},\boldsymbol{Z}) = E(\boldsymbol{b}) = \mathbf{0}$. Assumption (7) has the implication that teacher assignment for student i does not depend on the quality of the teacher (as measured by the b_g).

From an econometric perspective, equation (8) means that γ can be estimated in an unbiased way by OLS regression of

$$y_i \text{ on } \mathbf{x}_i, \;\; i = 1, ..., N. \tag{9}$$

Consequently, we can estimate the effects of the covariates $\boldsymbol{x_i}$ by omitting the teacher assignment dummies. Practically, this means we are assuming teacher assignment is uncorrelated with the covariates $\boldsymbol{x_i}$.

Under (5) and (7), the OLS estimator of γ is unbiased and consistent, but it is inefficient if we impose the standard classical linear model assumptions on $\boldsymbol{u}$. In particular, if

$$Var(\mathbf{u}|\mathbf{X}, \mathbf{Z}, \mathbf{b}) = Var(\mathbf{u}) = \sigma^2\mathbf{I}_N \tag{10}$$

then

$$
\begin{aligned}
Var(\mathbf{y}|\mathbf{X},\mathbf{Z}) &= E[(\mathbf{Zb}+\mathbf{u})(\mathbf{Zb}+\mathbf{u})'|\mathbf{X},\mathbf{Z}] \\
&= \mathbf{Z}Var(\mathbf{b})\mathbf{Z}' + Var(\mathbf{u}) = \sigma_b^2\mathbf{ZZ}' + \sigma_u^2\mathbf{I}_N,
\end{aligned}
$$

where we also add the standard assumption that the elements of $\boldsymbol{b}$ are uncorrelated

$$
Var(\mathbf{b}) = \sigma_b^2\mathbf{I}_G, \tag{11}
$$

and σ_b^2 is the variance of the teacher effects, b_g.

Under the assumption that σ_b^2 and σ_u^2 are known – actually, it suffices to know their ratio – the BLUE of γ under the preceding assumptions is the generalized least squares (GLS) estimator,

$$
\gamma^* = [\mathbf{X}'(\sigma_b^2\mathbf{ZZ}' + \sigma_u^2\mathbf{I}_N)^{-1}\mathbf{X}]^{-1}\mathbf{X}'(\sigma_b^2\mathbf{ZZ}' + \sigma_u^2\mathbf{I}_N)^{-1}\mathbf{y}. \tag{12}
$$

The $N \times N$ matrix $\boldsymbol{ZZ}'$ is a block diagonal matrix with G blocks, where block g is an $N_g \times N_g$ matrix of ones and N_g is the number of students taught by teacher g. The GLS estimator γ^* is the well-known "random effects" (RE) estimator popular from panel data and cluster sample analysis. Note that the "random effects" in this case are teacher effects, not student-specific effects.

Before we discuss γ^* further, as well as estimation of $\boldsymbol{b}$, it is helpful to write down the mixed effects model in perhaps a more common form. After students have been designated to classrooms, we can write y_{gi} as the outcome for student i in class g, and similarly for $\boldsymbol{x_{gi}}$ and u_{gi}. Then, for classroom g, we have

$$y_{gi} = \mathbf{x}_{gi}\gamma + b_g + u_{gi} \equiv \mathbf{x}_{gi}\gamma + r_{gi}, \;\; i = 1, ..., N_g, \tag{13}$$

where $r_{gi} \equiv b_g + u_{gi}$ is the composite error term. Equation (13) makes it easy to see that the BLUE of γ is the random effects estimator. It also highlights the assumption that b_g is independent of the covariates $\boldsymbol{x_{gi}}$. Further, the assumption $E(u_{gi}|\boldsymbol{X_g}, b_g) = 0$ implies that covariates from student h do not affect the outcome of student i. We can also see that OLS pooled across i and g is unbiased for γ because we are assuming $E(b_g|\boldsymbol{X_g}) = 0$.

As shown in, say, BSW, the BLUP of $\boldsymbol{b}$ under assumptions (5), (7), and (10) is

$$\mathbf{b}^* = (\mathbf{Z}'\mathbf{Z} + \rho\mathbf{I}_G)^{-1}\mathbf{Z}'(\mathbf{y} - \mathbf{X}\gamma^*) \equiv (\mathbf{Z}'\mathbf{Z} + \rho\mathbf{I}_G)^{-1}\mathbf{Z}'\mathbf{r}^*, \tag{14}$$

where $\rho = \sigma_u^2/\sigma_b^2$ and $\boldsymbol{r}^* = \boldsymbol{y} - \boldsymbol{X}\gamma^*$ is the vector of residuals. Straightforward matrix algebra shows each b_g^* can be expressed as

$$\begin{aligned} b_g^* &= (N_g + \rho)^{-1} \sum_{i=1}^{N_g} r_{gi}^* = \left(\frac{N_g}{N_g + \rho}\right) \bar{r}_g^* \\ &= \left(\frac{\sigma_b^2}{\sigma_b^2 + (\sigma_u^2/N_g)}\right) \bar{r}_g^* = \left(\frac{\sigma_b^2}{\sigma_b^2 + (\sigma_u^2/N_g)}\right) (\bar{y}_g - \bar{\mathbf{x}}_g\gamma^*), \end{aligned} \tag{15}$$

where

$$\bar{r}_g^* = N_g^{-1} \sum_{i=1}^{N_g} r_{gi}^* = \bar{y}_g - \bar{\mathbf{x}}_g \gamma^* \tag{16}$$

is the average of the residuals $r_{gi}^* = y_{gi} - \boldsymbol{x_{gi}}\gamma^*$ within classroom g.

To operationalize γ^* and b_g^* we must replace σ_b^2 and σ_u^2 with estimates. There are different ways to obtain estimates depending on whether one uses OLS residuals after an initial estimation or a joint estimation method. With the composite error defined as $r_{gi} = b_g + u_{gi}$ we can write $\sigma_r^2 = \sigma_b^2 + \sigma_u^2$. An estimator of σ_r^2 can be obtained from the usual sum of squared residuals from the OLS regression

$$y_{gi} \text{ on } \mathbf{x}_{gi},\ i = 1, ..., N_g,\ g = 1, ..., G. \tag{17}$$

Call the residuals $\widetilde{r_{gi}}$. Then a consistent estimator is

$$\tilde{\sigma}_r^2 = \frac{1}{(N-K)} \sum_{g=1}^{G} \sum_{i=1}^{N_g} \tilde{r}_{gi}^2, \tag{18}$$

which is just the usual degrees-of-freedom (df) adjusted error variance estimator from OLS.

To estimate σ_u^2, write

$$r_{gi} - \bar{r}_g = u_{gi} - \bar{u}_g,$$

where $\overline{r_g}$ is the within-teacher average, and similarly for $\overline{u_g}$. A standard result on demeaning a set of uncorrelated random variables with the same variance gives $Var(u_{gi} - \overline{u_g}) = \sigma_u^2(1 - N_g^{-1})$ and so, for each g, $E[\sum_{i=1}^{N_g}(r_{gi} - \overline{r_g})^2] = \sigma_u^2(N_g - 1)$. When we sum across teachers it follows that

$$\frac{1}{(N-G)}\sum_{g=1}^{G}\sum_{i=1}^{N_g}(r_{gi} - \bar{r}_g)^2 \tag{19}$$

has expected value σ_u^2. To turn (19) into an estimator we can replace the r_{gi} with the OLS residuals, as before, $\widetilde{r_{gi}}$, from the regression in (17). The estimator based on the OLS residuals is

$$\tilde{\sigma}_u^2 = \frac{1}{(N-G)}\sum_{g=1}^{G}\sum_{i=1}^{N_g}(\tilde{r}_{gi} - \bar{\tilde{r}}_g)^2. \tag{20}$$

With fixed class sizes and G getting large, the estimator that uses N in place of $N - G$ is not consistent. Therefore, we prefer the estimator in equation (20), as it should have less bias in applications where G/N is not small. With many students per teacher the difference should be minor. We could also use $N - G - K$ as a further df adjustment, but subtracting off K does not affect the consistency.

Given $\widetilde{\sigma_r^2}$ and $\widetilde{\sigma_u^2}$ we can estimate σ_b^2 as

$$\tilde{\sigma}_b^2 = \tilde{\sigma}_r^2 - \tilde{\sigma}_u^2. \tag{21}$$

In any particular data set – especially if the data have been generated to violate the standard assumptions listed above – there is no guarantee that expression (21) is nonnegative. A simple solution to this problem (and one used in is software packages, such as Stata) is to set $\widetilde{\sigma_b^2} = 0$ whenever $\widetilde{\sigma_r^2} < \widetilde{\sigma_u^2}$. In order to ensure this happens infrequently with multiple cohorts, we compute $\widetilde{\sigma_u^2}$ by replacing $\overline{\widetilde{r_g}}$ with the average obtained for the particular cohort. This ensures that, for a given cohort, the terms $\sum_{i=1}^{N_g}\left(\widetilde{r_{gi}} - \overline{\widetilde{r_g}}\right)^2$ are as small as possible. In theory, if there are no cohort effects we could use an overall cohort mean for $\overline{\widetilde{r_g}}$. But using cohort-specific means reduces the problem of negative $\widetilde{\sigma_b^2}$ when the model is misspecified.

An appealing alternative is to estimate σ_b^2 and σ_u^2 jointly along with γ, using software that ensures nonnegativity of the variance estimates. The most common approach to doing so is to assume joint normality of the teacher effects, b_g, and the student effects, u_{gi}, across all g and i – along with the previous assumptions. One important point is that the resulting estimators are consistent even without the normality assumption; so, technically, we can think of them as "quasi-" maximum likelihood estimators. The maximum likelihood estimator of σ_u^2 has the same form as in equation (20), except the residuals are based on the MLE of γ rather than the OLS estimator. A similar comment holds for the MLE of σ_b^2 (if we do not constrain it to be nonnegative). See, for example, Hsiao (2003, Section 3.3.3).

Unlike the GLS estimator of γ, the feasible GLS (FGLS) estimator is no longer unbiased (even under assumptions (5) and (7)), and so we must rely on asymptotic theory. In the current context, the estimator is consistent and asymptotically normal provided $G \to \infty$ with N_g fixed.[24]

[24] In simulations, Hansen (2007) shows that the asymptotic properties work well when G and N_g are roughly around 40.

In practice, this means that the number of teachers, G, should be substantially larger than the number of students per teacher, N_g. Typically this is the case in VAM studies, which are applied to large school districts or entire states and therefore include many teachers. Often the number of students per teacher is fewer than 100 with several hundred or even several thousand teachers.

When γ^* is replaced with the FGLS estimator and the variances σ_b^2 and σ_u^2 are replaced with estimators, the EB estimator of $\boldsymbol{b}$ is no longer a BLUP. Nevertheless, we use the same formula as in (15) for operationalizing the BLUPs. Conveniently, certain statistical packages – such as Stata with its "xtmixed" command – allow one to recover the operationalized BLUPs after maximum likelihood estimation. When we use the (quasi-) MLEs to obtain the b_g^* we obtain what are typically called the empirical Bayes' estimates.

One way to understand the shrinkage nature of b_g^* is to compare it with the estimator obtained by treating the teacher effects as fixed parameters. Let $\hat{\gamma}$ and $\hat{\beta}$ be the OLS estimators from the regression

$$\mathbf{y}_i \text{ on } \mathbf{x}_i,\ \mathbf{z}_i,\ i = 1, ..., N. \tag{22}$$

Then $\hat{\gamma}$ is the so-called "fixed effects" (FE) estimator obtained by a regression of y_i on the controls in $\boldsymbol{x_i}$ and the teacher assignment dummies in $\boldsymbol{z_i}$. In the context of the model

$$y = \mathbf{X}\gamma + \mathbf{Z}\beta + \mathbf{u}$$

$$E(\mathbf{u}|\mathbf{X},\mathbf{Z}) = \mathbf{0},\ Var(\mathbf{u}|\mathbf{X},\mathbf{Z}) = \sigma_u^2\mathbf{I}_N, \tag{23}$$

$\hat{\gamma}$ is the BLUE of γ and $\hat{\beta}$ is the BLUE of β. As is well-known, $\hat{\gamma}$can be obtained by an OLS regression where y_{gi} and $\boldsymbol{x_{gi}}$ have been deviated from within-teacher averages (see, for example, Wooldridge 2010, Chapter 10). Further, the estimated teacher fixed effects can be obtained as

$$\hat{\beta}_g = \bar{y}_g - \bar{\mathbf{x}}_g \hat{\gamma}. \tag{24}$$

Equation (24) makes computation of the teacher VAMs fairly efficient if one does not want to run the long regression in (22).

By comparing equations (15) and (24) we see that the EB estimator b_g^* differs from the fixed effects estimator $\widehat{\beta_g}$ in two ways. First, and most importantly, the RE estimator γ^* is used in computing b_g^* while $\widehat{\beta_g}$ uses the FE estimator $\hat{\gamma}$. Second, b_g^* shrinks the average of the residuals toward zero by the factor

$$\frac{\sigma_b^2}{\sigma_b^2 + (\sigma_u^2/N_g)} = \frac{1}{1 + (\rho/N_g)} \tag{25}$$

where

$$\rho = \sigma_u^2/\sigma_b^2. \tag{26}$$

Equation (25) illustrates the well-known result that the smaller is the number of students for teacher g, N_g, the more the average residual is shrunk toward zero.

A well-known algebraic result – for example, Wooldridge (2010, Chapter 10) – that holds for any given number of teachers G is that

$$\gamma^* \to \hat{\gamma} \text{ as } \rho \to 0 \text{ or } N_g \to \infty.^{25} \tag{27}$$

Equation (27) can be verified by noting that the RE estimator of γ can be obtained from the pooled OLS regression

$$y_{gi} - \theta_g \bar{y}_g \text{ on } \mathbf{x}_{gi} - \theta_g \; \bar{\mathbf{x}}_g \tag{28}$$

where

$$\theta_g = 1 - \left(\frac{\sigma_u^2}{\sigma_u^2 + N_g \sigma_b^2}\right)^{1/2} = 1 - \left(\frac{1}{1 + (N_g/\rho)}\right)^{1/2}. \tag{29}$$

It is easily seen that $\theta_g \to 1$ as $\rho \to 0$ or $N_g \to \infty$. In other words, with many students per teacher or large teacher effects relative to student effects, the RE and FE estimates can be very close. But they are never identical. Not coincidentally, the shrinkage factor in equation (25) tends to unity if $\rho \to 0$ or $N_g \to \infty$. The bottom line is that with a "large" number of students per teacher the shrinkage estimates of the teacher effects can be close to the fixed effects estimates. The RE and

[25] Lockwood and McCaffrey (2007) have highlighted equation (27) in the context of student-level panel data, essentially appealing to the first edition of Wooldridge (2010). In the panel data setting (27) is arguably less relevant, as one rarely has more than a handful of time periods per student. For additional discussion of the relationship between random and fixed effects estimators, see Raudenbush (2009). In addition, Reardon and Raudenbush (2009) lay out the various assumptions underlying value-added estimation.

FE estimates also tend to be similar when σ_u^2 (the student effect) is "small" relative to σ_b^2 (the teacher effect), but this scenario seems unlikely.

An important point that appears to go unnoticed in applying the shrinkage approach is that in situations where γ^* and $\hat{\gamma}$ substantively differ, γ^* suffers from systematic bias because it assumes teacher assignment is uncorrelated with $\boldsymbol{x_i}$. Because γ^* is used in constructing the b_g^* in equation (15), the bias in γ^* generally results in biased teacher effects, and the teacher effects would be biased even if (15) did not employ a shrinkage factor. The shrinkage likely exacerbates the problem: the estimates are being shrunk toward values that are systematically biased for the teacher effects.[26]

The expression in equation (15) motivates a common two-step alternative to the EB approach proper. In the first step of the procedure, one obtains $\tilde{\gamma}$ using the OLS regression in equation (17), and obtains the residuals, $\widetilde{r_{gi}}$. In the second step, one averages the residuals $\widetilde{r_{gi}}$ within each teacher to obtain the teacher effect for teacher g. We call this approach the "average residual" (AR) method. After obtaining the averages of the residuals one can, in a third step, shrink the averages using the empirical Bayes' shrinkage factors in equation (15). Typically the estimates in equations (18) and (20), based on the OLS residuals, are used in obtaining the shrinkage factors. We call the resulting estimator the "shrunken average residual" (SAR) method.

[26] Without covariates, the difference between the EB and fixed effects estimates of the b_g is much less important: they differ only due to the shrinkage factor. In practice, the fixed effects estimates, $\widehat{\beta_g}$, are obtained without removing an overall teacher average, which means $\widehat{\beta_g} = \overline{y_g}$. To obtain a comparable expression for b_g^* we must account for the GLS estimator of the mean teacher effect, which would be obtained as the intercept (the only parameter) in the RE estimation. Call this estimator μ_b^*, which in the case of no covariates is γ^*. Then the teacher effects are

$$b_g^* = \mu_b^* + \eta_g(\bar{y}_g - \mu_b^*) = \eta_g \bar{y}_g + (1 - \eta_g)\mu_b^* = \bar{y}_g - (1 - \eta_g)(\bar{y}_g - \mu_b^*),$$

where η_g is the shrinkage factor in equation (25). Compared with the FE estimate of b_g, b_g^* is shrunk toward the overall mean μ_b^*. When the teacher effects are treated as parameters to estimate, the b_g^* are biased because of the shrinkage factor, even in the case in which they are BLUP.

With or without shrinking, the AR approach suffers from systematic bias if teacher assignment, $\boldsymbol{z_i}$, is correlated with the covariates, $\boldsymbol{x_i}$. In effect, the AR approach partials $\boldsymbol{x_i}$ out of y_i but does not partial $\boldsymbol{x_i}$ out of $\boldsymbol{z_i}$, the latter of which is crucial if $\boldsymbol{x_i}$ and $\boldsymbol{z_i}$ are correlated. The so-called "fixed effects" regression in (22) partials $\boldsymbol{x_i}$ out of $\boldsymbol{z_i}$, which makes it a more reliable estimator under nonrandom teacher assignment – perhaps much more reliable with strong forms of nonrandom assignment.

It is also important to know that the SAR approach is inferior to the EB approach under nonrandom assignment. The logic is simple. First, the algebraic relationship between RE and FE means that γ^* tends to be closer to the FE estimator, $\hat{\gamma}$, than the OLS estimator, $\tilde{\gamma}$. Consequently, under nonrandom teacher assignment, the estimated teacher effects using the RE estimator of γ will have less bias than the estimates that begin with OLS estimation of γ. Second, if teacher assignment *is* uncorrelated with the covariates, the OLS estimator of γ is inefficient relative to the RE estimator under the standard random effects assumptions (because the RE estimator is FGLS). Thus, the only possible justification for SAR is computational simplicity. But the saving is likely to be minor unless the number of controls in $\boldsymbol{x_i}$ is very large. For the kinds of data sets widely available, the computational saving from using SAR rather than EB is likely to be minor.

Before we leave this section we must emphasize that fixed effects estimation of the teacher VAMs allows any correlation between $\boldsymbol{z_i}$ and $\boldsymbol{x_i}$, and thus expect it to outperform EB estimation and strongly outperform SAR under nonrandom assignment. The bias due to nonrandom allocation of students to teachers is also discussed in Rothstein (2009, 2010).

3.3 Summary of Estimation Methods

In this paper we examine five different value-added estimators used to recover the teacher effects and apply them to both real and simulated data. Some of the estimators use EB or shrinkage techniques, while others do not. They can all be cast as a special case of the estimators described in the previous section. For clarity, we briefly describe each one, with additional reference to each of these specifications provided in Table 18.

The estimators can be obtained from a dynamic equation of the form

$$A_{it} = \lambda A_{i,t-1} + \mathbf{X}_{it}\delta + \mathbf{Z}_{it}\beta + v_{it}, \tag{30}$$

in which A_{it} is achievement (measured by a test score) for student i in grade t, $\boldsymbol{X}_{it}$ is a vector of student characteristics, and $\boldsymbol{Z}_{it}$ is the vector of teacher assignment dummies. This is similar to equation (1) but with the lagged test score written separately from $\boldsymbol{X}_{it}$ for clarity. Also, $\boldsymbol{X}_{it}$ is omitted from the estimation of the teacher effects in the simulation analysis below, as student characteristics are not included in the data generating process. The EB estimator we analyzed in Section 3.2 was for the case of a single cross-section of students. Thus, we use only one grade – fifth grade – for the analysis.

We first analyze EB LAG, a dynamic MLE version of the EB estimator that treats the teacher effects as random. This technique obtains the estimates of the teacher effects using the normal maximum likelihood in the first stage, regressing A_{it} on its lag, $A_{i,t-1}$, and $\boldsymbol{X}_{it}$. In the second stage, the shrinkage factor is applied to these teacher effects.[27] A second estimator we consider is the average residual (AR) method described in Section 3.2. This technique mainly

[27] As described in Rabe-Hesketh and Skrondal (2012), this two step procedure can be performed in one-step using the "xtmixed" command in Stata with teacher random effects. The predicted random effects of this regression are identical to shrinking the MLE estimates by the shrinkage factor. This procedure is generally justified even if the unobservables do not have normal distributions, in which case we are applying quasi-MLE.

differs from EB LAG in that it uses OLS in the first stage. The residuals of this OLS regression are obtained, and then we average these residuals by classroom to calculate the estimated teacher effects. We expect the EB LAG estimator to outperform the AR estimator in most scenarios, given that MLE is being used in the first-stage instead of OLS.

We compare the estimators that treat the teacher effect as random with an estimator that explicitly controls for the teacher effect through the inclusion of teacher assignment dummy variables. This third estimator applies OLS to (30) by pooling across students and classrooms. We refer to this estimator as "dynamic OLS," or DOLS. DOLS treats the teacher effects as fixed parameters to estimate. The inclusion of the lagged test score accounts for the possibility that teacher assignment is related to the past test score. Guarino, Reckase, and Wooldridge (forthcoming) discuss the assumptions under which DOLS consistently estimates β when (30) is obtained from a structural cumulative effects model, and the assumptions are quite restrictive. Nevertheless, their simulations show the DOLS estimator often estimates β well even when the assumptions underlying the consistency of DOLS fails.

Given that EB estimation is often motivated as a way to increase precision and decrease misclassification, we also analyze whether shrinking AR and DOLS estimates enhances performance. Thus, the fourth estimator we analyze, SAR (for shrunken average residual), takes the AR estimates and shrinks them by the shrinkage factor described in Section 3.2 using the variance estimates from equations (18), (20), and (21). Shrinking the AR estimates does not result in a true EB estimator since AR uses OLS in the first stage, but it is commonly used as a simpler way of operationalizing the EB approach.[28] As discussed in Section 3.2, with a

[28] See, for example, Kane and Staiger, 2008). AR and SAR will be fairly similar with large class sizes and will be consistent under the same assumptions. The finite-sample performance of these estimators will differ only due to the shrinkage factor. It is important to keep in mind that, unlike DOLS and SDOLS, the AR and SAR estimators do not allow for general correlation between the teacher assignment and past test scores (or other covariates).

sufficiently large number of students per teacher, the EB LAG estimator converges to the DOLS estimator but SAR does not. Thus, as the number of students per teacher grows, we would expect EB LAG to perform more similarly to DOLS than SAR. Finally, we consider a shrunken DOLS (SDOLS) estimator, which takes the DOLS estimated teacher fixed effects and shrinks them by the shrinkage factor derived in Section 3.2. Although SDOLS is rarely done in practice and is not a true EB estimator, we include it as an exploratory exercise in order to better determine the effects of shrinking itself when the number of students per teacher differs. When the class sizes are all the same, the SDOLS and DOLS estimates differ only by a constant positive multiple and shrinking the DOLS estimates will have no effect in terms of ranking teachers.

3.4 Comparing VAM Methods Using Simulated Data

The question of which VAM estimators perform the best can only be addressed in simulations in which the true teacher effects are known. Therefore, to evaluate the performance of EB estimators relative to other common value-added estimators, we apply these methods to simulated data. This approach allows us to examine how well various estimators recover the true teacher effect under a variety of assignment scenarios. Using data generated from the processes described in Section 3.4.1, we apply the set of value-added estimators discussed in Section 3.3 and compare the resulting estimates with the true underlying teacher effects.

3.4.1 Simulation Design

Much of our main analysis focuses on a base case that restricts the data generating process to a relatively narrow set of idealized conditions. These ideal conditions do not allow for measurement error or peer effects and assume that teacher effects are constant over time. The data are constructed to represent grades three through five (the tested grades) in a hypothetical school. For simplicity and comparison with the theoretical predictions, we assume that the learning process has been going on for a few years but only calculate estimates of teacher effects for fifth grade teachers – a single cross section.[29] We create data sets that contain students nested within teachers nested within schools, with students followed longitudinally over time in order to reflect the institutional structure of an elementary school. Our simple baseline data generating process is as follows:

$$
\begin{aligned}
A_{i3} &= \lambda A_{i2} + \beta_{i3} + c_i + e_{i3} \\
A_{i4} &= \lambda A_{i3} + \beta_{i4} + c_i + e_{i4} \\
A_{i5} &= \lambda A_{i4} + \beta_{i5} + c_i + e_{i5}
\end{aligned} \tag{32}
$$

in which A_{i2} is a baseline score reflecting the subject-specific knowledge of child i entering third grade, A_{it} is the grade-t test score ($t = 3,4,5$), λ is a time constant decay parameter,[30] β_{it} is the teacher-specific contribution to growth (the true teacher value-added effect), c_i is a time-invariant student-specific effect (may be thought of as "ability" or "motivation"), and e_{it} is a random deviation for each student.[31] In all of the simulations reported in this paper, the random

[29] Despite only estimating value-added for grade 5 teachers, we keep the three grade structure when generating the student test scores to more realistically capture how fifth grade test scores are determined. The fifth grade achievement is based on more than just the current teacher and prior test score of the student; it is a function of all prior teacher, unobservable student, and random influences. Thus, to ignore that process and generate fifth grade test scores based on a "baseline" fourth grade test score seems inappropriate given this context.

[30] For lag scores greater than one year prior, lambda is set equal to zero in the simulations. Some models, however, use multiple prior test scores (e.g. EVAAS, VARC) and we estimate DOLS, AR, and EB LAG with multiple lagged test scores as a sensitivity analysis. Although adding multiple lags improves the performance of AR and EB LAG in the random assignment case, the performance of these estimators still suffers greatly compared to DOLS in the DG-PA and DG-NA scenarios. Thus, the results of this sensitivity analysis suggest that adding additional lags to equation (30) does little to change our overall conclusions.

[31] Because we assume independence of e_{it} over time we are maintaining the so-called "common factor restriction" in the underlying cumulative effects model. This restriction implies that past shocks to student learning decay at the

variables A_{i2}, c_i, and e_{it} are drawn from normal distributions with means of zero. The true teacher effect, β_{it}, is drawn from a normal distribution with a mean of 0.5. The standard deviation of the teacher effect is .25,[32] the standard deviation of the student fixed effect is .5, and the standard deviation of the random noise component is 1. These give relative shares of 5, 19, and 76 percent of the total variance in gain scores (when $\lambda = 1$), respectively. Given that the student and noise components are larger than the teacher effects, we call these "small" teacher effects.[33] The baseline score is drawn from a distribution with a standard deviation of 1. We also allow for correlation between the time-invariant child-specific heterogeneity, c_i, and the baseline test score, A_{i2}, which we set to 0.5. This correlation reflects that students with better unobserved "ability" likely have higher test scores as well.

Our data are simulated using 36 teachers and 720 students per cohort. In order to create a situation in which there is a substantial variation in class size – to showcase the potential disparities between EB/shrinkage and other estimators – we vary the number of students per classroom. Teachers receive classes of varying sizes, but receive the same number of students in each cohort. Of the 36 teachers we simulate, nine teachers have classes of 10 students, nine teachers have a class size of 20, and nine teachers have class sizes of 30. We simulate the data using both one and four cohorts of students to provide further variance in the amount of data from which the teacher effects are calculated. In the case of four cohorts, data are pooled across the cohorts so that value-added estimates are based off of sample sizes of 40, 80, and 120,

same rate as all inputs. See Guarino, Reckase, and Wooldridge (2012) for a more detailed discussion of this assumption.

[32] The mean and standard deviation of the true teacher effects of the 36 teachers we estimate are 0.501 and 0.244, respectively.

[33] We also conduct a sensitivity analysis using "large" teacher effects, where the true teacher effects are drawn from a distribution with a standard deviation of 1. When teacher effects are large, the performance of all estimators is increased and EB LAG performs similarly to DOLS in the DG-PA and DG-NA cases. The AR method, however, continues to suffer in performance under DG-PA and DG-NA compared to DOLS.

instead of 10, 20, and 30 as in the one cohort case. Therefore, we would expect that estimates in the four cohort case to be less noisy than those from the one cohort case, possibly mitigating the potential gains from EB estimation.

To create different scenarios, we vary certain key features: the grouping of students into classes, the assignment of classes of students to teachers within schools, and the level of persistence in prior learning from one year to the next. We generate data using each of the nine different mechanisms for the assignment of students outlined in Table 23. Students are grouped into classrooms either randomly, based on their prior year achievement level (dynamic grouping or DG), or based on their unobserved heterogeneity (heterogeneity grouping or HG). In the random case, students are assigned a random number and then grouped into classrooms of various sizes based on that random number. In the grouping cases, students are ranked by either the prior test score or the student fixed effect and grouped into classrooms of various sizes based on that ranking. Teachers are assigned to these classrooms either randomly (denoted RA) or nonrandomly. Teachers assigned nonrandomly can be assigned positively (denoted PA), meaning the best teachers are assigned to classrooms with the best students, or negatively (denoted NA), meaning the best teachers are assigned to classrooms with the worst students.

These grouping and assignment procedures are not purely deterministic, as we allow for a a random component with standard deviation of one in the assignment mechanism.[34] We use the estimators discussed in Section 3.3, but with only a constant, teacher dummies (if applicable), and, for the dynamic specifications, the lagged test score included as covariates. We use 100 Monte Carlo replications per scenario in evaluating each estimator.

[34] As a sensitivity analysis, we also run simulations with this standard deviation = 0.1, meaning the grouping of students into classrooms is more nonrandom. In this case, the performance of AR and EB LAG suffers even more greatly in terms of lower rank correlations and higher misclassification rates than what is observed in the results in Tables18 and 19.

3.4.2 Evaluation Measures

For each estimator across each iteration, we save the individual estimated teacher effects and also retain the true teacher effects, which are fixed across the iterations for each teacher. To study how well the methods uncover the true teacher effects, we adopt five simple summary measures using the teacher level data. The first is a measure of how well the estimates preserve the rankings of the true teacher effects. We compute the Spearman rank correlation, $\hat{\rho}$, between the estimated teacher effects and the true effects and report the average $\hat{\rho}$ across the 100 iterations. Second, we compute a measure of misclassification. These misclassification rates are obtained as the percentage of above average teachers (in the true quality distribution) who are misclassified as below average in the distribution of estimated teacher effects for the given estimator.

In addition to examining rank correlations and misclassification rates, it is also helpful to have a measure that quantifies some notion of the magnitude of the bias in the estimates. Given that some teacher effects are biased upwards and others downwards, it is difficult to capture the overall bias in the estimates in a simple way. We create a statistic, $\hat{\theta}$, that captures how closely the magnitude of the deviation of the estimates from their mean tracks the size of the deviation of the true effects from the true mean. To calculate this measure, we regress the deviation of the estimated teacher effects from their overall estimated means on the analogous deviation of the true effects generated from the simulation for each estimator. We can represent this simple regression as

$$\hat{\beta}_j - \bar{\hat{\beta}} = \hat{\theta}(\beta_j - \bar{\beta}) + residual_j, \tag{33}$$

in which $\widehat{\beta_j}$ is the estimated teacher effect and β_j is the true effect of teacher j. From this simple regression, we report the average coefficient, $\bar{\hat{\theta}}$, across the 100 replications of the simulation for each estimator. This regression tells us whether the estimated teacher effects are correctly distributed around the average teacher. If $\hat{\theta} = 1$, then a movement of β_j away from its mean is tracked by the same movement of $\widehat{\beta_j}$ from its mean.

When $\hat{\theta} \approx 1$, the magnitudes of the estimated teacher effects can be compared across teachers. If $\hat{\theta} > 1$, the estimated teacher effects amplify the true teacher effects. In other words, teachers above average will be estimated to be even more above average and vice versa for below average teachers. An estimation method that produces $\hat{\theta}$ substantially above one generally does a good job of ranking teachers, but the magnitudes of the differences in estimated teacher effects cannot be trusted. The magnitudes also cannot be trusted if $\hat{\theta} < 1$; in this case, ranking the teachers becomes more difficult because the estimated effects are compressed relative to the true teacher effects.

In addition to ranking teachers correctly, the magnitude of the estimated teacher effects is also important in policy applications. It is helpful to examine the extent to which shrinking the estimates, as in the EB methods, increases bias in these noisy estimates. Thus, we report the average value of $\hat{\theta}$ across the simulations because it provides evidence of which methods, under which scenarios, produce estimated teacher effects whose magnitudes have meaning. This measure also provides insight into why some methods rank teachers relatively well even when the estimated effects are systematically biased.

The precision of these methods is also a key consideration when evaluating the overall performance. As described in Section 3.2, EB methods are not unbiased when thinking about the teacher effects as fixed parameters we are trying to estimate. However, if the identifying assumptions hold, these methods should provide more precise estimates. This is one motivation for using such methods, as estimates should be more stable over time, leading to a smaller variance in the teacher effects. As the teacher effect is fixed for each teacher across the 100 iterations, we have 100 estimates of each teacher effect. As a summary measure for the precision of the estimators, we calculate the standard deviation of the 100 teacher effect estimates for each teacher and then take a simple average across all teachers.

To further analyze the variance-bias tradeoff for each of these estimators, we also include average mean squared error (MSE). This measure averages the following across all j teachers and across simulation runs:

$$\widehat{\mathrm{MSE}} = (\beta_j - \hat{\beta}_j)^2 \tag{34}$$

This provides a simple statistic to determine whether the bias induced by shrinking is justifiable due to gains in precision.

3.5 Simulation Results

Tables 18 and 19 report the five evaluation measures described in Section 3.4.2 for each particular estimator-assignment scenario combination. For ease in interpreting the tables, a quick guide to descriptions of each of these estimators, grouping-assignment mechanisms, and evaluation measures can be found in the appendix in Tables 23 through 25. As these shrinkage and EB estimators are often motivated as a way to reduce noise, one might expect these approaches to be most beneficial with very limited student data per teacher. Thus, we estimate teacher effects using both four cohorts and one cohort of data. The tables show results for the case $\lambda = .5$. Though not reported, we also conducted a full set of simulations for $\lambda = 0.75$ and $\lambda = 1$, and the main conclusions are unchanged. The full set of simulation results is available upon request from the authors.

3.5.1 Fixed Teacher Effects versus Random Teacher Effects

We first compare the performance of the DOLS estimator, which treats teacher effects as fixed parameters to estimate, to the AR and EB LAG estimators that treat teacher effects as random in Table 18. Under nonrandom assignment of teachers, we expect DOLS, which explicitly controls for teacher assignment through the inclusion of teacher assignment indicators, to perform better than those estimators treating the teacher effects as random. When teacher assignment is based on the lagged test score, DOLS directly controls for the assignment mechanism by including both the lagged score and teacher assignment indicators and should perform particularly well in this case. The simulation results presented here largely support this hypothesis.

3.5.1.1 Random Assignment

We begin with the pure random assignment (RA) case, where EB-type estimation methods are theoretically justified. The results of the random assignment case are given in the top panel of Table 18 and suggest that the difference between fixed and random effects estimators is not that substantial under this scenario. As the theory suggests, EB LAG performs well in the four cohort case, with rank correlations between the estimated and the true teacher effects near 0.76, slightly better than the 0.75 rank correlation for DOLS. The AR estimator, which uses OLS in the first stage instead of MLE, outperforms both DOLS and EB LAG in terms of the rank

TABLE 18
Simulation Results: Comparing Fixed and Random Teacher Effects Estimators

$\lambda = 0.5$		Four Cohorts			One Cohort		
G-A Mechanism	Evaluation Type	**DOLS**	**AR**	**EB LAG**	**DOLS**	**AR**	**EB LAG**
RA	Rank Correlation	0.75	0.78	0.76	0.61	0.62	0.63
	Misclassification	0.21	0.19	0.21	0.28	0.27	0.28
	Avg. Theta	1.01	1.00	0.82	1.03	1.02	0.53
	Avg. Std. Dev.	0.28	0.27	0.23	0.38	0.37	0.20
	MSE	0.15	0.28	0.14	0.20	0.34	0.16
DG-RA	Rank Correlation	0.75	0.77	0.76	0.59	0.60	0.61
	Misclassification	0.22	0.20	0.22	0.29	0.28	0.29
	Avg. Theta	1.00	0.98	0.80	0.99	0.97	0.50
	Avg. Std. Dev.	0.28	0.28	0.23	0.37	0.36	0.19
	MSE	0.15	0.29	0.14	0.20	0.33	0.16
DG-PA	Rank Correlation	0.76	0.60	0.70	0.58	0.45	0.48
	Misclassification	0.23	0.34	0.28	0.30	0.36	0.37
	Avg. Theta	0.99	0.75	0.70	0.96	0.73	0.38
	Avg. Std. Dev.	0.28	0.23	0.21	0.37	0.33	0.18
	MSE	0.15	0.33	0.14	0.21	0.38	0.15
DG-NA	Rank Correlation	0.74	0.54	0.68	0.60	0.36	0.43
	Misclassification	0.22	0.31	0.25	0.28	0.37	0.49
	Avg. Theta	1.01	0.46	0.51	1.03	0.48	0.15
	Avg. Std. Dev.	0.28	0.18	0.17	0.38	0.30	0.12
	MSE	0.14	0.31	0.20	0.20	0.36	0.26
HG-RA	Rank Correlation	0.67	0.67	0.68	0.56	0.56	0.57
	Misclassification	0.26	0.26	0.26	0.29	0.29	0.30
	Avg. Theta	1.02	1.01	0.87	0.99	0.98	0.57
	Avg. Std. Dev.	0.32	0.32	0.28	0.40	0.39	0.24
	MSE	0.17	0.33	0.16	0.22	0.38	0.17
HG-PA	Rank Correlation	0.89	0.88	0.89	0.78	0.77	0.78
	Misclassification	0.13	0.14	0.14	0.20	0.21	0.22
	Avg. Theta	1.61	1.57	1.45	1.60	1.55	1.10
	Avg. Std. Dev.	0.41	0.40	0.37	0.48	0.46	0.34
	MSE	0.17	0.36	0.15	0.22	0.41	0.16
HG-NA	Rank Correlation	0.32	0.29	0.31	0.25	0.21	0.24
	Misclassification	0.39	0.39	0.39	0.40	0.42	0.55
	Avg. Theta	0.33	0.27	0.18	0.34	0.28	0.07
	Avg. Std. Dev.	0.18	0.17	0.11	0.29	0.28	0.10
	MSE	0.20	0.32	0.21	0.25	0.36	0.22

Note: Rows of each scenario represent the following:
First - Rank corr. of estimated effects and true effects
Second - Fraction of above avg. teachers misclassified below avg.; Third - Avg. value of $\hat{\theta}$
Fourth - Average standard deviation of estimated teacher effects across 100 reps
Fifth - MSE measure

correlation even though it is not theoretically preferred. In addition to very similar rank correlations, the misclassification rates are very similar across the three estimators, with between 19 to 21 percent of above average teachers misclassified as below average. The similarities between the three estimators in terms of rank correlation and misclassification rates remain when using only one cohort. Reducing the amount of data used to estimate the teacher effects lowers the performance of all estimators, decreasing the rank correlations and increasing the misclassification rates. With one cohort, rank correlations between the estimated and true teacher effects are about 0.61 to 0.63, and between 27 and 28 percent of above average teachers are misclassified as below average.

In addition to rank correlations and misclassification rates, we also examine the bias and precision of the estimators. While DOLS and AR appear to be unbiased with average $\hat{\theta}$ values close to 1, EB LAG substantially underestimates the magnitudes of the true teacher effects with an average $\hat{\theta}$ value of 0.82 using four cohorts and 0.53 using one cohort. This bias is likely the result of the shrinkage technique that is applied. However, this shrinkage causes EB LAG to be slightly more precise than AR and DOLS. While DOLS and AR both have similar average standard deviations of the estimated teacher effects near 0.28 and 0.38 in the four and one cohort cases, respectively, EB LAG has lower average standard deviations of 0.23 and 0.20, respectively. Given the precision gain in EB LAG, the MSE measure suggests that EB LAG may be preferred to DOLS under random assignment. The MSE for AR suggests that even under random assignment, DOLS and EB LAG would be preferred.

We now move to the cases where the students are *nonrandomly grouped* together, but teachers are still *randomly assigned* to classrooms, the DG-RA and HG-RA panels in Table 18.

Under these two scenarios, we see a fairly similar pattern as in the RA scenario, although the overall performance of all estimators is slightly diminished.

3.5.1.2 Dynamic Grouping and Nonrandom Assignment

The performance of the various estimators diverges noticeably under nonrandom teacher assignment. We allow for nonrandom grouping based on either the prior year test score or student-level heterogeneity, but now allow for nonrandom assignment of students to teachers. Classes with high test scores or high unobserved ability can be assigned to either the best (positive assignment) or worst (negative assignment) teachers. A key finding of this analysis is the disparity in performance of estimators that treat teacher effects as random (e.g. AR and EB LAG) compared with the DOLS estimator. These results suggest that estimators explicitly controlling for the teacher assignment should be preferred to those that treat the teacher effects as random.

DOLS substantially outperforms AR and EB LAG under the DG-PA scenario. When using four cohorts, DOLS has a rank correlation of 0.76 under DG-PA, while AR and EB LAG have rank correlations of 0.60 and 0.70, respectively. AR and EB LAG also have large misclassification rates, with 28 to 34 percent of above average teachers being misclassified as below average compared with only 23 percent for DOLS.[35] In addition to misclassifying and poorly ranking teachers, the AR and EB LAG methods also underestimate the magnitudes of the true teacher effects. While DOLS has an average $\hat{\theta}$ value of 0.99, the AR and EB LAG

[35] Although the results are omitted from the paper we also examine the fraction of above average teachers that are misclassified in the bottom quintile and the fraction of below average teachers that are misclassified in the top quintile. As expected, the conclusions are very similar to those drawn from the misclassification rates reported in the tables. Teachers are more likely to be misclassified in the extremes by AR and EB LAG under DG-PA and DG-NA than under random assignment, while DOLS misclassifies teachers at similar rates across all scenarios. Shrinking the estimates also does not appear to have much impact on these misclassification rates.

estimators have average $\hat{\theta}$ values of 0.75 and 0.70, respectively. While some of the bias of the EB LAG estimates can be attributed to shrinkage, the larger issue is the bias caused by the failure of the AR and EB LAG approaches to net out the correlation between the lagged test score and the teacher assignment (i.e. the assignment mechanism in these DG scenarios), a correlation that DOLS explicitly allows for with the inclusion of teacher dummies in the regression. Just as in the random assignment case, DOLS and EB LAG have similar MSE measures, while the MSE for AR is substantially larger. In the four cohort case, DOLS, EB LAG and AR have MSE values of 0.15, 0.14, and 0.29, respectively.

These differences are even more noticeable under the DG-NA scenario. Again examining the four cohort case, DOLS has a rank correlation of 0.74, while AR and EB LAG have rank correlations of 0.54 and 0.68, respectively. Only 22 percent of above average teachers are misclassified by DOLS, while 31 and 25 percent are misclassified by AR and EB LAG, respectively. AR and EB LAG are even more biased under DG-NA, with $\hat{\theta}$ values of 0.46 and 0.51, respectively. This severe underestimation again offsets the gain in precision of the AR and EB LAG estimators, leading to MSE measures that are larger than for DOLS.

These simulation results also verify an important result of the theoretical discussion: the performance of EB LAG approaches the performance of DOLS as the number of students per teacher grows. We see less of a disparity in the performance of DOLS and EB LAG when computing VAMs using four cohorts compared to one, but the relative performance of AR does not improve with more students per teacher. For example, under DG-PA with one cohort of students, AR and EB LAG have similar rank correlations of 0.45 and 0.48, respectively, compared to 0.58 for DOLS. With four cohorts of students, the rank correlation for EB LAG is much closer to that for DOLS (0.70 and 0.76, respectively) than is the rank correlation for AR

(0.60). This theoretical result is also applicable to the SAR estimator we examine below, which is used as a simpler way to operationalize the EB approach. In summary, EB LAG, which uses random effects estimation in the first stage, is preferred to those using OLS (AR and SAR) under nonrandom teacher assignment, as these estimates approach the preferred DOLS estimates that treat teacher effects as fixed.

3.5.1.3 Heterogeneity Grouping and Nonrandom Assignment

As a final scenario we examine the case of nonrandom teacher assignment to students grouped on the basis of student-level heterogeneity. The results for these HG scenarios are especially unstable: all estimators do an excellent job ranking teachers under positive teacher assignment, and all estimators do a very poor job under negative teacher assignment. In the HG-PA case with four cohorts of students, the magnitudes of the estimated VAMs are amplified as seen by the large average values for $\hat{\theta}$ between 1.45 and 1.61. This improves the ability of the various estimators to rank teachers as evidenced by the high rank correlations of about 0.89 for all estimators. The EB LAG estimator performs the best in this scenario, as it performs as well as the other estimators in terms of ranking and misclassification of teachers but has the smallest MSE measure. Under HG-NA with four cohorts, the performance of all estimators falls substantially, largely caused by severely underestimated teacher effects ($\hat{\theta}$ values between 0.18 and 0.33). These compressed teacher effect estimates make it difficult to rank teachers in this scenario, resulting in low rank correlations for all estimators between 0.29 and 0.32. Just as in the HG-PA scenario, the performance of the three estimators under HG-NA is very similar across the evaluation measures we examine.

Why is the performance of DOLS, AR, and EB LAG so similar under HG-PA and HG-NA, while differing so greatly under DG-PA and DG-NA? Despite correlation between the baseline test score and the student fixed effect, the lagged test score appears to be a weak proxy for the assignment mechanism in the HG scenarios. Since none of the three estimators do well at allowing for the correlation between the assignment mechanism and the teacher assignment in these cases, the distinction between estimators that include teacher fixed effects and those that treat teacher effects as random is less stark. As found in Guarino, Reckase, and Wooldridge (forthcoming), a gain score estimator with student fixed effects included is the most robust in these HG scenarios, as it does allow for the correlation between the assignment mechanism (i.e. student fixed effect) and the teacher assignment (i.e. teacher dummy variables). Their results lend further support for our conclusions here that allowing for this correlation is extremely important in the performance of these value added estimators when there is nonrandom assignment.

3.5.2 Shrinkage versus Non-Shrinkage Estimation

Use of EB and other shrinkage estimators is often motivated as a way to reduce the noise in the estimation of teacher effects, particularly for teachers with a small number of students. Greater stability in the estimated effects is thought to reduce misclassification of teachers. We observed in Section 3.3.1 that EB LAG was generally outperformed by the fixed effects estimator, DOLS. However, under nonrandom teacher assignment, we are unable to tell how much of the bias in the EB LAG estimator is due to treating the teacher effects as random and how much is due to the shrinkage procedure. To examine the effects of shrinkage itself, we compare the performance of unshrunken estimators, DOLS and AR, with their shrunken versions, SDOLS and SAR, in Table 19. Although SDOLS is not a commonly used or theoretically justified estimator, we explore it here to identify whether shrinking teacher fixed effect estimates could be useful in practice.

Our simulation results show that there is no substantial improvement in the performance of the DOLS or AR estimators after applying the shrinkage factor to the estimates. Using four cohorts of students, the performance measures for DOLS and AR compared to their shrunken counterparts are nearly identical to two decimal places across all grouping and assignment scenarios. Even with very limited data per teacher in the one cohort case, when we would expect shrinkage to have a greater effect on the estimates, we find very little change in the performance of the estimators after the shrinkage factor is applied.

In the one cohort case, shrinking either the DOLS or AR estimates slightly decreases (in the second decimal place) both the average $\hat{\theta}$ values and average standard deviation of the estimated teacher effects. This increased bias in the estimates is expected when applying the shrinkage factor and, depending on the scenario and estimator we examine, the effect of this precision-bias tradeoff may increase or decrease the MSE measure when comparing the shrunken and unshrunken estimates. Shrinking the DOLS estimates generally reduces both the misclassification of teachers and the MSE measure slightly. Shrinking the AR estimates doesn't affect misclassification in most cases, and it actually increases misclassification slightly in the DG-NA scenario. Effects on the MSE measure are mixed for shrinking the AR estimates, but generally reduce the MSE measure slightly.

The effect of shrinkage itself does not appear to be practically important or ameliorate the performance of the biased AR estimator found in the DG-PA and DG-NA scenarios. Given that shrinking the AR estimates does little to mitigate the performance drop of AR under DG-PA and

DG-NA, our evidence suggests that shrinking teacher fixed effects estimates is preferred over shrinking teacher random effects if such techniques are desired.

TABLE 19
Simulation Results: Comparing Shrunken and Unshrunken Estimators

$\lambda = 0.5$		Four Cohorts				One Cohort			
G-A Mechanism	Evaluation Type	**DOLS**	**SDOLS**	**AR**	**SAR**	**DOLS**	**SDOLS**	**AR**	**SAR**
	Rank Correlation	0.75	0.75	0.78	0.78	0.61	0.61	0.62	0.62
	Misclassification	0.21	0.21	0.19	0.19	0.28	0.26	0.27	0.27
RA	Avg. Theta	1.01	1.01	1.00	1.00	1.03	0.99	1.02	0.98
	Avg. Std. Dev.	0.28	0.28	0.27	0.27	0.38	0.36	0.37	0.35
	MSE	0.15	0.15	0.28	0.28	0.20	0.18	0.34	0.33
	Rank Correlation	0.75	0.75	0.77	0.77	0.59	0.58	0.60	0.60
	Misclassification	0.22	0.22	0.20	0.20	0.29	0.28	0.28	0.28
DG-RA	Avg. Theta	1.00	1.00	0.98	0.98	0.99	0.96	0.97	0.94
	Avg. Std. Dev.	0.28	0.28	0.28	0.28	0.37	0.36	0.36	0.34
	MSE	0.15	0.14	0.29	0.29	0.20	0.17	0.33	0.33
	Rank Correlation	0.76	0.76	0.60	0.60	0.58	0.59	0.45	0.45
	Misclassification	0.23	0.23	0.34	0.34	0.30	0.30	0.36	0.36
DG-PA	Avg. Theta	0.99	0.99	0.75	0.75	0.96	0.95	0.73	0.70
	Avg. Std. Dev.	0.28	0.28	0.23	0.23	0.37	0.36	0.33	0.31
	MSE	0.15	0.15	0.33	0.33	0.21	0.18	0.38	0.37
	Rank Correlation	0.74	0.74	0.54	0.54	0.60	0.61	0.36	0.36
	Misclassification	0.22	0.22	0.31	0.31	0.28	0.28	0.37	0.38
DG-NA	Avg. Theta	1.01	1.01	0.46	0.46	1.03	1.01	0.48	0.45
	Avg. Std. Dev.	0.28	0.28	0.18	0.18	0.38	0.36	0.30	0.27
	MSE	0.14	0.14	0.31	0.31	0.20	0.16	0.36	0.35
	Rank Correlation	0.67	0.67	0.67	0.67	0.56	0.56	0.56	0.56
	Misclassification	0.26	0.26	0.26	0.26	0.29	0.29	0.29	0.29
HG-RA	Avg. Theta	1.02	1.02	1.01	1.01	0.99	0.96	0.98	0.95
	Avg. Std. Dev.	0.32	0.32	0.32	0.32	0.40	0.39	0.39	0.38
	MSE	0.17	0.17	0.33	0.33	0.22	0.20	0.38	0.37
	Rank Correlation	0.89	0.89	0.88	0.88	0.78	0.78	0.77	0.77
	Misclassification	0.13	0.13	0.14	0.14	0.20	0.20	0.21	0.21
HG-PA	Avg. Theta	1.61	1.61	1.57	1.57	1.60	1.57	1.55	1.52
	Avg. Std. Dev.	0.41	0.41	0.40	0.40	0.48	0.47	0.46	0.45
	MSE	0.17	0.16	0.36	0.35	0.22	0.20	0.41	0.39
	Rank Correlation	0.32	0.32	0.29	0.29	0.25	0.27	0.21	0.22
	Misclassification	0.39	0.39	0.39	0.39	0.40	0.39	0.42	0.42
HG-NA	Avg. Theta	0.33	0.33	0.27	0.27	0.34	0.37	0.28	0.27
	Avg. Std. Dev.	0.18	0.18	0.17	0.17	0.29	0.27	0.28	0.26
	MSE	0.20	0.20	0.32	0.32	0.25	0.20	0.36	0.35

Note: Rows of each scenario represent the following: First - Rank corr. of estimated effects and true effects Second - Fraction of above average teachers misclassified as below average; Third - Average value of $\hat{\theta}$ Fourth - Average standard deviation of estimated teacher effects across 100 reps; Fifth - MSE measure

3.6 Comparing VAM Methods Using Real Data

We also apply these estimation methods to actual student-level test score data and examine the rank correlations between the estimated teacher effects of the various estimators for each school district. In addition to rank correlations, we also examine whether teachers are being classified in the extremes uniformly across all of the estimators we examine. Although the real data does not allow comparison between the estimated effects and the true teacher effects, we are able to make comparisons between the estimated effects of different estimators. This comparison provides a measure of the sensitivity of the estimated teacher effects to specifications that shrink the estimates and/or treat the teacher effects as random or fixed. The results of this analysis provide some perspective on the impact of shrinking and Empirical Bayes' methods in a real-world setting.

3.6.1 Data

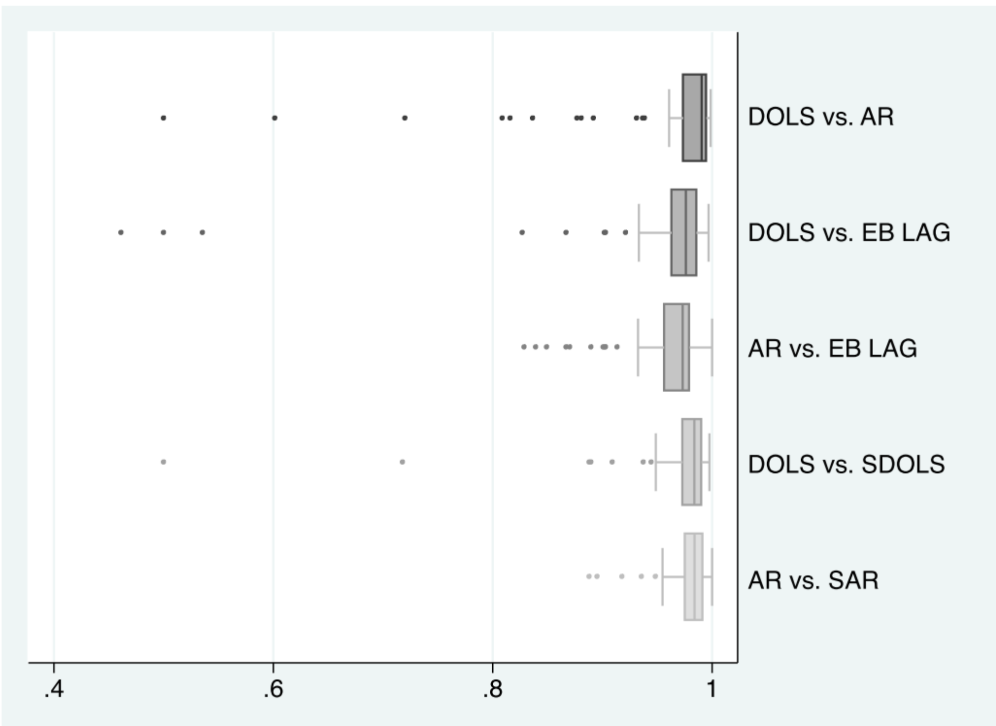
We apply the five methods described in Section 3.3 to data from an anonymous southern U.S. state. The data span 2001 through 2007 and grades four through six, but test scores are collected for each student from grades three through six. The data set includes 1,488,253 total students from which we have at least one current year score and one lagged score. Only 482,031 students have test scores for all grades. The data set also contains 43,868 unique teachers that we observe for a varying number of cohorts of students. We observe 39 percent of teachers for only one year, but we do see 20 percent of teachers for four or more years. These teachers, on average, teach about 26 students per year, with only a small percentage (less than two percent) teaching more than 30 students per year. The high percentage of teachers that we observe for only one year could motivate researchers to employ shrinkage and EB estimators as a way to reduce precision problems due to minimal data.

We estimate teacher effects district-by-district using equation (30) with controls for various student characteristics and include dummies for the year. Student characteristics include race, gender, disability status, free/reduced price lunch eligibility, limited English proficiency status, and the number of student absences from school. As discussed above, the teacher effects are estimated using data on multiple cohorts (between one and seven) of students. For simplicity and comparison with the simulation results, we estimate the value-added measures for those teachers with fifth grade students in the 67 districts, but again teachers receive multiple cohorts of students. Overall, we estimate 20,749 teacher effects using test score data from the annual assessment exam administered by the state.

3.6.2 Results

Figure 12 presents box plots that depict the distributions of the within-district rank correlations between the various lagged score estimators, DOLS, SDOLS, AR, SAR, and EB LAG. As in the discussion of the simulation results, we first compare the DOLS estimator, which treats the teacher effects as fixed, with the estimators that treat the teacher effects as random. Comparing DOLS and AR, we find that the median rank correlation is around 0.99, but there are nine districts with rank correlations below 0.90 and 2 districts with correlations below 0.50. We see a slightly lower median rank correlation between DOLS and EB LAG, at around 0.97, with five districts with rank correlations below 0.90 and 3 below 0.50. These results are not inconsistent with our simulation results: the performance of DOLS, AR, and EB LAG is very similar under cases of random assignment of teachers to classrooms, but the performance of AR and EB LAG is substantially different from DOLS under non-random assignment based on prior test scores. Thus, it could be the case that the outlier districts we see in the left tails of the top two box plots may be composed of schools that engage more heavily in nonrandom assignment of teachers to classrooms.

FIGURE 12
Spearman Rank Correlations Across Different VAM Estimators



Comparing the two estimators that treat teacher effects as random, AR and EB LAG, we find that while the median rank correlation is 0.96, nine districts have rank correlations of between 0.82 and 0.92. These results suggest that the estimates are somewhat sensitive to how the teacher effects are calculated in the first stage. This was also the case in the simulated results, where the performance of the AR estimator suffered more than the performance of the EB LAG estimator in cases of non-random assignment based on the prior test score.

For a thorough comparison with the simulation results, we also compare the shrunken and unshrunken estimates of DOLS and AR using the real data. We find median rank correlations of around 0.97 for both the DOLS and SDOLS comparison and the AR and SAR comparison, suggesting that shrinkage has a small impact on the estimates. Shrinkage may have a slightly larger impact on the DOLS estimates, as two districts have rank correlations of 0.50 and 0.72. Our simulation results suggested that shrinking the estimates had very little impact on estimator performance, but the SDOLS estimator showed the greatest boost in performance from shrinking, especially in the case of one cohort of students.

In addition to rank correlation comparisons, we also examine the extent to which teachers are classified in the tails of the distribution by the different estimators. If shrinkage is having some effect, we would expect to see some teachers classified in the extremes to be pushed toward the middle of the distribution after applying the shrinkage factor. Table 20 lists the fraction of teachers ranked in the same quintile, either the top or bottom, by different pairs of estimators. Comparing across estimators that assume fixed teacher effects to those that assume random teacher effects, we do not see much movement across quintiles. For example, comparing DOLS to EB LAG, we find that about 91 percent of the teachers that are classified in the top quintile using DOLS are also in this quintile using EB LAG. This suggests that teacher assignment may not be largely based on prior student achievement or that the prior test score is a poor proxy for the true assignment mechanism. If the prior test score or other covariates insufficiently proxy for the underlying assignment mechanism, then the choice to include teacher assignment variables will matter little in how teachers are ranked.

Comparing the rankings of unshrunken and corresponding shrunken estimators, we see that about 90 percent of teachers are ranked in the same quintile by both the unshrunken estimators (DOLS and AR) and their shrunken counterparts (SDOLS and SAR). This suggests that shrinking the estimates results in some reclassification of teachers in the tails to quintiles in the middle of the distribution. Using real data, however, we are unable to tell whether this reclassification is appropriate. Our simulated analysis suggested that shrinking the estimates had little impact if any on misclassification rates.

TABLE 20
Fraction of Teachers Ranked in Same Quintile by Estimator Pairs

	DOLS	SDOLS	AR	SAR
		Top Quintile		
SDOLS	0.91			
AR	0.94	0.89		
SAR	0.89	0.94	0.91	
EB LAG	0.87	0.95	0.86	0.93
		Bottom Quintile		
SDOLS	0.89			
AR	0.96	0.88		
SAR	0.88	0.95	0.89	
EB LAG	0.87	0.98	0.86	0.96

3.7 Conclusion

Using simulation experiments where the true teacher effects are known, we have explored the properties of two commonly used Empirical Bayes' estimators as well as the effects of shrinking estimates of teacher effects in general. Overall, EB methods do not appear to have much advantage, if any, over simple methods such as DOLS that treat the teacher effects as fixed, even in the case of random teacher assignment where EB estimation is theoretically justified. Under random assignment, all estimators perform well in terms of their ability of ranking teachers, properly classifying teachers, and providing unbiased estimates. EB methods have a very slight gain in precision compared to the other methods in this case.

We generally find that EB estimation is not appropriate under nonrandom teacher assignment. The hallmark of EB estimation of teacher effects is to treat the teacher effects as random variables that are independent (or at least uncorrelated) with any other covariates. This assumption is tantamount to assuming that teacher assignment does not depend on other covariates such as past test scores (this is also true for the AR methods). When teacher assignment is not random, estimators that either explicitly control for the assignment mechanism or proxy for it in some way typically provide more reliable estimates of the teacher effects. Among the estimators and assignment scenarios we study, DOLS and SDOLS are the only estimators that control for the assignment mechanism (again, either explicitly or by proxy) through the inclusion of both the lagged test score and teacher assignment dummies. As expected, DOLS and SDOLS outperform the other estimators in the nonrandom teacher assignment scenarios. In the analysis of the real data, we found that the rank correlations between, say, DOLS and EB LAG or DOLS and SAR are quite low for some districts, suggesting that the decision among these estimators is important. Thus, if there is a possibility of nonrandom assignment, DOLS should be the preferred estimator.

As predicted by theory and seen in the simulation results, the random effects estimator, EB LAG, converges to the fixed effects estimator, DOLS, as the number of students per teacher gets large. Therefore, it could be that EB LAG is performing well in large samples simply because the estimates are approaching the DOLS estimates. However, the average residual methods, AR and SAR, do not have this property. Thus, despite the recent popularity, we strongly caution using SAR as a simpler way to operationalize the EB LAG estimator. If EB-type methods are being used, it is important to estimate the coefficients in the first stage using random effects estimation (as in our EB LAG estimator) rather than OLS.

Lastly, we find that shrinking the estimates of the teacher effects does not seem to improve the performance of the estimators, even in the case where estimates are based on one cohort of students. The performance measures are extremely close in our simulations for those estimators that differ only due to the shrinkage factor – DOLS and SDOLS or AR and SAR. The rank correlations for these two pairs of estimators are also very close to one in almost all districts. Also, we find in the simulations that shrinking the AR estimates, which is a popular way to operationalize the EB approach, doesn't reduce misclassification of teachers. Thus, our evidence suggests that the rationale for using shrinkage estimators to reduce the misclassification of teachers due to noisy estimates of teacher effects should not be given much weight. Accounting for nonrandom teacher assignment when choosing among estimators is more imperative.

Given the robust nature of the DOLS estimator to a wide variety of grouping and assignment scenarios, it should be widely preferred to AR and EB methods when there is uncertainty about the true underlying assignment mechanism. If the assignment mechanism is known to be random, applying these AR and EB estimators can be appropriate, especially when the amount of data per teacher is minimal. However, given that the assignment mechanism is not likely known, blindly applying these AR and EB methods can be extremely problematic, especially if teachers are truly assigned nonrandomly to classrooms. Therefore, we stress caution in applying theses AR and EB methods and urge researchers to be mindful of the underlying assignment mechanism when choosing between the various value-added methods.

APPENDIX

APPENDIX

TABLE 21
Summary Statistics by Treatment and Unconditional Difference-in-Differences Estimates
Single Mothers w/ High School Education or Less

	One Child		Two Plus Children		
VARIABLES	1990-93	1995-99	1990-93	1995-99	Diff-in-Diff
In Poverty	0.417	0.372	0.644	0.582	-0.017
	(0.008)	(0.009)	(0.007)	(0.008)	(0.016)
In Labor Force	0.705	0.820	0.564	0.737	0.058***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.014)
Family Earned Income	16.180	18.560	10.620	13.970	0.970
	(0.283)	(0.384)	(0.199)	(0.340)	(0.619)
EITC Payment	0.668	1.307	0.586	2.059	0.834***
	(0.012)	(0.023)	(0.011)	(0.031)	(0.042)
Welfare + Food Stamps	2.047	1.178	5.462	3.433	-1.160***
	(0.057)	(0.049)	(0.084)	(0.077)	(0.137)
Family Net Income	21.740	23.280	21.020	23.390	0.830
	(0.257)	(0.332)	(0.204)	(0.294)	(0.552)
Observations	5,023	3,601	6,909	4,989	

Notes: Weighted by CPS household weights. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 22
Main Difference-in-Differences Results
Single Mothers w/ High School Education or Less

VARIABLES	In Poverty	In Labor Force	Earned Income	EITC Payment	Welfare + FS	Net Income
*After * Two Plus*	0.000	0.051***	0.546	0.841***	-1.090***	0.303
	(0.014)	(0.016)	(0.605)	(0.055)	(0.108)	(0.594)
Two Plus Children	0.139***	-0.100***	-3.901***	-0.041*	2.544***	1.974***
	(0.009)	(0.008)	(0.330)	(0.022)	(0.119)	(0.337)
Welfare Max Benefit	-0.015**	-0.024***	0.627**	0.048**	0.053	1.038**
	(0.007)	(0.005)	(0.283)	(0.024)	(0.122)	(0.400)
Time Limits on Welfare	-0.007	-0.010	-0.463	0.021	-0.134	-0.699
	(0.018)	(0.018)	(0.737)	(0.083)	(0.218)	(0.686)
Black	0.047	0.055	1.266	0.069	0.450	-0.939
	(0.035)	(0.047)	(1.051)	(0.086)	(0.300)	(1.226)
White	-0.088**	0.109**	4.247***	0.110	-0.944***	2.715**
	(0.034)	(0.044)	(1.058)	(0.084)	(0.272)	(1.182)
Asian	-0.091**	0.085	3.805***	0.228	-0.809*	3.417*
	(0.045)	(0.066)	(0.880)	(0.163)	(0.415)	(1.954)
Children Under Five	0.148***	-0.154***	-2.951***	-0.221***	1.700***	-2.490***
	(0.009)	(0.010)	(0.465)	(0.028)	(0.226)	(0.322)
Age	-0.024***	0.024***	1.713***	0.010	0.002	0.749***
	(0.003)	(0.003)	(0.101)	(0.008)	(0.024)	(0.101)
Age^2	0.000***	-0.000***	-0.020***	-0.000**	-0.001**	-0.003**
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
Year Fixed Effects	x	x	x	x	x	x
State Fixed Effects	x	x	x	x	x	x
Highest Grade FEs	x	x	x	x	x	x
Observations	20,521	20,521	20,521	20,521	20,521	20,521
R-squared	0.243	0.171	0.168	0.225	0.278	0.171

Notes: Robust standard errors (clustered at the state level) in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Each column represents a separate WLS regression using CPS household weights for the various outcomes.

TABLE 23
Definitions of Grouping-Assignment Mechanisms

Name of G-A Mechanism	**Acronym**	**Grouping students in classrooms**	**Assigning students to teachers**
Random Assignment	**RA**	Random	Random
Dynamic Grouping - Random Assignment	**DG-RA**	Dynamic (based on prior test scores)	Random
Dynamic Grouping - Positive Assignment	**DG-PA**	Dynamic (based on prior test scores)	Positive corr. between teacher effects and prior student scores
Dynamic Grouping - Negative Assignment	**DG-NA**	Dynamic (based on prior test scores)	Negative corr. between teacher effects and prior student scores
Heterogeneity Grouping - Random Assignment	**HG-RA**	Static (based on student heterogeneity)	Random
Heterogeneity Grouping - Positive Assignment	**HG-PA**	Static (based on student heterogeneity)	Positive corr. between teacher effects and student fixed effects
Heterogeneity Grouping - Negative Assignment	**HG-NA**	Static (based on student heterogeneity)	Negative corr. between teacher effects and student fixed effects

TABLE 24
Description of Evaluation Measures of Value-Added Estimator Performance

Evaluation Measure	Description
Rank Correlation	Rank correlation between estimated teacher effect and true teacher effect
Misclassification	Fraction of above average teachers that are misclassified as below average
Average Theta	Average value of $\hat{\theta}$
Avg. Std. Dev.	Average standard deviation of estimated teacher effects across the 100 simulation reps
MSE	Average value of $\widehat{MSE} = (\beta_j - \hat{\beta}_j)^2$

TABLE 25
Description of Value-Added Estimators

Estimator	Acronym	Description	Teacher Effects
Empirical Bayes'	EB LAG	Two-step approach: Estimate teacher effects using MLE on dynamic equation and then shrink estimates by shrinkage factor	Random
Average Residual	AR	Estimate dynamic equation by OLS and compute residuals for each student. Then compute the average of these residuals for each teacher to get estimated teacher effect	Random
Dynamic OLS	DOLS	Estimate teacher effects using ordinary least squares on dynamic equation	Fixed
Shrunken DOLS	SDOLS	Two-step approach: Estimate teacher effects using dynamic equation and then shrink estimates by shrinkage factor	Fixed
Shrunken Avg. Residual	SAR	Two-step approach: Compute average residual for each teacher using residuals from OLS on dynamic equation. Then shrink average residual for each teacher by shrinkage factor	Random

BIBLIOGRAPHY

BIBLIOGRAPHY

Adireksomdat, K. (2010). "The Effects of the 1993 Earned Income Tax Credit Expansion on the Labor Supply of Unmarried Women," Public Finance Review 38(1): 11-40.

Baker, K. (2008). "Do Cash Transfer Programs Improve Infant Health: Evidence from the 1993 Expansion of the Earned Income Tax Credit," mimeo, University of Notre Dame.

Ballou, D., Sanders, W., and Wright, P. (2004). "Controlling for Student Background in Value-Added Assessment of Teachers," Journal of Educational and Behavioral Statistics 29(1): 37-65.

Baughman, R. and Dickert-Conlin, S. (2003). "Did Expanding the EITC Promote Motherhood?" American Economic Review Papers and Proceedings 93(2): 247-250.

Baughman, R. and Dickert-Conlin, S. (2009). "The Earned Income Tax Credit and Fertility," Journal of Population Economics 22(3): 537-563.

Baughman, R. and Duchovny, N. (2012). "State EITCs and Production of Child Health: Insurance Coverage, Utilization, and Health Status," Working Paper.

Blank, R. M. (2002). "Evaluating Welfare Reform in the United States," NBER Working Paper 8983.

Bollinger, C. R. (1998). "Measurement Error in the Current Population Survey: A Nonparametric Look," Journal of Labor Economics 16(3): 576–94.

Bollinger, C. R. and David, M. H. (1997). "Modeling Discrete Choice with Response Error: Food Stamp Participation," Journal of the American Statistical Association 92(439): 827-35.

Bollinger, C. R. and David, M. H. (2005). "I Didn't Tell, and I Won't Tell: Dynamic Response Error in the SIPP," Journal of Applied Econometrics 20: 563-9.

Bound, J. and Krueger, A.B. (1991). "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" Journal of Labor Economics 9(1): 1–24.

Brooks-Gunn, J., Chatterji, P., and Markowitz, S. (2011). "Early Maternal Employment and Family Wellbeing," NBER Working Paper 17212.

Center for Human Resource Research (2004). "NLSY79 Child & Young Adult Data Users Guide," The Ohio State University, Columbus, OH.

Chetty, R., Friedman, J. N., and Rockoff, J. (2011a). "New Evidence on the Long-Term Impacts of Tax Credits," IRS Statistics of Income White Paper, 2011

Chetty, R., Freidman, J., and Rockoff, J. (2011b). "The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood," NBER Working Paper 17699.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Whitmore Schanzenbach, D., and Yagan, D. (2011c). "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star," The Quarterly Journal of Economics 126 (4): 1593-1660.

Chetty, R. and Saez, E. (2013). "Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients," forthcoming, American Economic Journal: Applied Economics.

Chetty, R., Friedman, J. N., and Saez, E. (2013). "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings," forthcoming, American Economic Review.

Corcoran, S., Jennings, J., and Beveridge, A. (2011). "Teacher effectiveness on high- and low-stakes tests," Unpublished Draft.

Dahl, M., DeLeire, T., and Schwabish, J. (2009). "Stepping Stone or Dead End? The Effect of the EITC on Earnings Growth," IZA Discussion Paper No. 4146.

Dahl, G. B. and Lochner, L. (2012). "The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit," American Economic Review 102(5): 1927-56.

Dickert-Conlin, S. and Houser, S. (2002). "EITC and Marriage," National Tax Journal 60 (1): 25–39.

Dickert, S., Houser, S., and Scholz, J. K. (1995). "The Earned Income Tax Credit and Transfer Programs: A Study of Labor Market and Program Participation," Tax Policy and the Economy 9: 1-50.

Duncan, G. J., Brooks-Gunn, J., Yeung, W. J., and Smith, J. R. (1998). "How Much Does Childhood Poverty Affect the Life Chances of Children?" American Sociological Review 63: 406-23.

Eissa, N. and Hoynes, H. W. (2000). "Good News for Low Income Parents: Tax-Transfer Schemes and Marriage," mimeo, University of California, Berkeley.

Eissa, N. and Hoynes, H. W. (2004). "Taxes and the Labor Market Participation of Married Couples: The Earned Income Tax Credit," Journal of Public Economics 88(9-10): 1931-58.

Eissa, N. and Hoynes, H. W. (2006). “Behavioral Responses to Taxes: Lessons from the EITC and Labor Supply,” Tax Policy and the Economy 20: 74-110.

Eissa, N. and Liebman, J. B. (1996). “Labor Supply Response to the Earned Income Tax Credit,” Quarterly Journal of Economics 111(2): 605-637.

Ellwood, D. T. (2000). “The Impact of the Earned Income Tax Credit and Social Policy Reforms on Work, Marriage, and Living Arrangements,” National Tax Journal 43(4, part 2): 1063–105.

Evans, W. and Garthwaite, C. (2014). “Giving Mom a Break: The Impact of Higher EITC Payments on Maternal Health,” American Economic Journal: Economic Policy 6(2): 258-90.

Feenberg, D. and Coutts, E. (1993). “An Introduction to the TAXSIM Model.” Journal of Policy Analysis and Management 12(1): 189–94.

Finn, J. D., Boyd-Zaharias, J., and Gerber, S. B. (2004). “Small Classes in the Early Grades, Academic Achievement, and Graduating from High School,” Journal of Educational Psychology 97: 214–23.

Friedman, P. (2000). "The Earned Income Tax Credit," Welfare Information Network, Issue Notes.

Goldberger, W. A., Prause, J., Lucas-Thompson, R., and Himsel, A. (2008). “Maternal Employment and Children’s Achievement in Context: A Meta-Analysis of Four Decades of Research,” Psychological Bulletin 134(1): 77-108.

Goodman-Bacon, A. and McGranahan, L. (2008). “How do EITC recipients spend their refunds?” Economic Perspectives 32(2): http://ssrn.com/abstract=1134060.

Grogger, J. (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 85(2): 394–408.

Guarino, C. M., Reckase, M. D., and Wooldridge, J. M. (2012). "Can Value-Added Measures of Teacher Performance Be Trusted?," Education Policy Center at Michigan State University Working Paper 18.

Hansen, C. B. (2007). "Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data when T is Large," Journal of Econometrics 141: 597-620.

Heckman, J. J. and Learner, E. (2001). Handbook of Econometrics Vol. 5. Elsevier.

Hill, J. L., Waldfogel, J., Brooks-Gunn, J., and Han, W. (2005). "Maternal Employment and Child Development: A Fresh Look Using Newer Methods," Developmental Psychology 41(6): 833-50.

Hotz, V. J. and Scholz, J. K. (2003). "The Earned Income Tax Credit," Robert Moffitt, ed., Means-Tested Transfer Programs in the United States. Chicago: University of Chicago Press.

Hotz, V. J., Mullin, C. H., and Scholz, J, K. (2006). "Examining the Effect of the Earned Income Tax Credit on the Labor Market Participation of Families on Welfare," NBER Working Paper 11968.

Hoynes, H. W., and Schanzenbach, D. W. (2009). "Consumption Reponses to In-Kind Transfers: Evidence from the Introduction of the Food Stamp Program," American Economic Journal: Applied Economics 1(4): 109-139.

Hoynes, H. W., Miller, D. L., and Simon, D. (2012). "Income, the Earned Income Tax, and Infant Health," Working Paper.

Internal Revenue Service (2002). "Participation in the Earned Income Tax Credit Program for Tax Year 1996," Washington, D.C.: IRS, January.

Internal Revenue Service (2011). Statistics of Income Branch, "SOI Tax Stats - Individual Statistical Tables by Size of Adjusted Gross Income, Individual Income Tax Returns with Earned Income Credit," http://www.irs.gov/pub/irs-soi/08in25ic.xls, accessed July 2011.

Jacob, B. and Lefgren, L. (2005), "Principals as Agents: Subjective Performance Measurement in Education," NBER Working Paper 11463.

Jacob, B. and Lefgren, L. (2008). "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education," Journal of Labor Economics 26(1): 101-136.

James-Burdumy, S. (2005). "The Effect of Maternal Labor Force Participation on Child Development," Journal of Labor Economics 23(1): 177-211.

Kane, T. and Staiger, D. O. (2008). "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation," NBER Working Paper 14607.

King, M., Ruggles, S., Alexander, J. T., Flood, S., Genadek, K., Schroeder, M. B., Trampe, B., and Vick, R. (2010). Integrated Public Use Microdata Series, Current Population Survey: Version 3.0. [Machine-readable database]. Minneapolis: University of Minnesota.

Krueger, A. B. (1999). "Experimental Estimates of Education Production Functions," Quarterly Journal of Economics 114(2): 497–532.

Krueger, A. B. and Whitmore, D. M. (2001). “The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR,” Economic Journal 111(January): 1–28.

LaLumia, S. (2013). “The EITC, Tax Refunds, and Unemployment Spells,” American Economic Journal: Economic Policy 5(2): 188-221.

Leigh, A. (2010). “Who Benefits from the Earned Income Tax Credit? Incidence among Recipients, Coworkers and Firms,” IZA Discussion Paper No. 4960.

Levine, P. P. and Zimmerman, D. J. (2005). “Children’s Welfare Exposure and Subsequent Development,” Journal of Public Economics 89: 31–56.

Liebman, J. B. (1998). “The Impact of the Earned Income Tax Credit on Incentives and Income Distribution,” Tax Policy and the Economy 12: 83-119.

Liebman, J. B. (2000). “Who are the Ineligible EITC Recipients?” National Tax Journal 53(4, part 2): 1165–85.

Lockwood, J. R. and McCaffrey, D. (2007). "Models for Value-Added Modeling of Teacher Effects," Electronic Journal of Statistics 1: 223-252.

Løken, K. V., Mogstad, M., and Wiswall, M. (2012). "What Linear Estimators Miss: The Effects of Family Income on Child Outcomes," American Economic Journal: Applied Economics 4(2): 1–35.

Maag, E. (2005). “Paying the Price? Low-Income Parents and the Use of Paid Tax Preparers,” New Federalism: National Survey of America’s Families B-64, Urban Institute.

Manoli, D. S. and Turner, N. (2014). “Cash-on-Hand & College Enrollment: Evidence from Population Tax Data and Policy Nonlinearities,” NBER Working Paper 19836.

Maxfield, M. (2013). “The Effects of the Earned Income Tax Credit on Child Achievement and Long-Term Educational Attainment,” Working Paper.

McCaffrey, D., Lockwood, J. R., Louis, T., and Hamilton, L. (2004). "Controlling for Individual Heterogeneity in Longitudinal Models, with Applications to Student Achievement," Journal of Educational and Behavioral Statistics 29(1): 67-101.

Meyer, B. D. (2007). "The U.S. Earned Income Tax Credit: Its Effects and Possible Reforms," Swedish Economic Policy Review 14(2): 55 – 80.

Meyer, B. D. and Rosenbaum, D. T. (1999). “Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers,” NBER Working Paper 7363.

Meyer, B. D. and Rosenbaum, D. T. (2000). “Making Single Mothers Work: Recent Tax and Welfare Policy and Its Effects,” National Tax Journal 53(4, Part 2): 1027–1061.

Meyer, B. D. and Rosenbaum, D. T. (2001). “Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers,” Quarterly Journal of Economics 116(3): 1063–1114.

Meyer, B. D. and Sullivan, J. X. (2004). “The Effects of Welfare and Tax Reform: The Material Well-Being of Single Mothers in the 1980s and 1990s.” Journal of Public Economics 88: 1387–1420.

Michelmore, K. (2013). “The Effect of Income on Educational Attainment: Evidence from State Earned Income Tax Credit Expansions,” Working Paper.

Milligan, K. and Stabile, M. (2011). “Do Child Tax Benefits Affect the Wellbeing of Children? Evidence from Canadian Child Benefit Expansions,” American Economic Journal: Economic Policy 3(3): 175-205.

Mink, P. T. (1993). “Children’s Initiatives in the Budget Reconciliation,” 103rd Cong., 1st session, Congressional Record, Vol. 139, July 29, p. H5511.

Morris, C. (1983). "Parametric Empirical Bayes Inference: Theory and Applications," Journal of the American Statistical Association 78(381): 47-55.

Morris, P. A., Huston, A. C., Duncan, G. J., Crosby, D. A., and Bos, J. M. (2001). “How Welfare and Work Policies Affect Children: A Synthesis of Research,” New York, NY: MDRC.

Morris, P. A., Duncan, G. J., and Clark-Kauffman, E. (2005). “Child Well-Being in an Era of Welfare Reform: The Sensitivity of Transitions in Development to Policy Change,” Developmental Psychology 41: 919–32.

National Commission on Children (1993). “Children’s Initiatives in the Budget Reconciliation,” 103rd Cong., 1st session, Congressional Record, Vol. 139, July 29, p. H5508.

Neumark, D. and Wascher, W. (2001). “Using the EITC to Help Poor Families: New Evidence and a Comparison with the Minimum Wage,” National Tax Journal 54(2): 281–317.

Olson, L. M. and Davis, A. (1994). “The Earned Income Tax Credit: Views from the Street Level,” Northwestern University Working Paper WP-94-1.

Rabe-Hesketh, S. and Skrondal, A. (2012). Multilevel and Longitudinal Modeling Using Stata, 3e. Stata Press: College Station, TX.

Raudenbush, S. W. (2009). "Adaptive centering with random effects: An alternative to the fixed effects model for studying time-varying treatments in school settings." Education Finance and Policy, 4(4): 468-491.

Reardon, S. F., & Raudenbush, S. W. (2009). "Assumptions of value-added models for estimating school effects." Education Finance and Policy 4(4): 492-519.

Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). “Teachers, Schools and Academic Achievement,” Econometrica 73: 417–458.

Rockoff, J. E. (2004). “The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data,” American Economics Review 94: 247–252.

Romich, J. and Weisner, T. (2000). ‘‘How Families View and Use the EITC Advance Payment versus Lump Sum Delivery,” National Tax Journal 53: 1245–66.

Ross-Phillips, K. (2001). “Who Knows About the Earned Income Tax Credit?” Urban Institute Policy Brief No. B-27.

Rothstein, J. (2007). “The Mid-1990s EITC Expansion: Aggregate Labor Supply Effects and Economic Incidence,” mimeo, Princeton University.

Rothstein, J. (2009). "Student sorting and bias in value-added estimation: Selection on observables and unobservables. Education Finance and Policy 4(4): 537-571.

Rothstein, J. (2010). “Teacher quality in educational production: Tracking, decay, and student achievement." The Quarterly Journal of Economics 125(1): 175-214.

Rowe, G. (2000). “State TANF Policies as of July 1999.” Report. Washington, DC: The Urban Institute.

Saez, E. (2010). “Do Taxpayers Bunch at Kink Points?” American Economic Journal: Economic Policy 2: 180–212.

Scholz, J. K. (1997). Testimony for the House Ways and Means Committee. May 8, Washington, D.C.

Shepard, K. (1993). “Children’s Initiatives in the Budget Reconciliation,” 103rd Cong., 1st session, Congressional Record, Vol. 139, July 29, p. H5512.

Shonkoff, J. P. and Phillips, D. (2000). From Neurons to Neighborhoods: The Science of Early Childhood Development. Committee on Integrating the Science of Early Childhood Development, Board on Children, Youth, and Families, Institute of Medicine, Division of Behavioral and Social Sciences and Education. Washington, D.C.: National Academy Press.

Smeeding, T. M., Ross-Phillips, K., and O’Connor, M. (2000). “The EITC: Expectation, Knowledge, Use, and Economic and Social Mobility,” National Tax Journal 53(4, part 2): 1187–209.

Smolensky, E. and Gootman, J. A. (2003). Working Families and Growing Kids: Caring for Children and Adolescents. Board on Children, Youth, and Families. Division of

Behavioral and Social Sciences and Education. Washington D.C.: National Academies Press.

Stupak, B. (1993). "Children's Initiatives in the Budget Reconciliation," 103rd Cong., 1st session, Congressional Record, Vol. 139, July 29, p. H5515.

Strully, K., Rehkopf, D. H., and Xuan, Z. (2010). "Effects of Prenatal Poverty on Infant Health: State Earned Income Tax Credits and Birth Weight," American Sociological Review 75(4): 534-62.

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

Ventry, D. J. (2000). "The Collision of Tax and Welfare Politics: the Political History of the Earned Income Tax Credit, 1969-1999," National Tax Journal 53(2): 983-1026.

Wooldridge, J. M. (2010). Econometric Analysis of Cross Section and Panel Data, 2e. MIT Press: Cambridge, MA.