Northwestern | INSTITUTE FOR POLICY RESEARCH

Working Paper Series

WP-17-09

# Reducing Inequality Through Dynamic Complementarity: Evidence from Head Start and Public School Spending

**Rucker C. Johnson**
Associate Professor
Goldman School of Public Policy
University of California, Berkeley

**C. Kirabo Jackson**
Associate Professor of Human Development and Social Policy
IPR Fellow
Northwestern University

Version: May 26, 2017

DRAFT

*Please do not quote or distribute without permission.*

## ABSTRACT

The researchers explore whether early childhood human-capital investments are complementary to those made later in life. Using the Panel Study of Income Dynamics, they compare the adult outcomes of cohorts who were differentially exposed to policy-induced changes in pre-school (Head Start) spending and school-finance-reform-induced changes in public K12 school spending during childhood, depending on place and year of birth. Difference-in-difference instrumental variables and sibling-difference estimates indicate that, for poor children, increases in Head Start spending and increases in public K12 spending each individually increased educational attainment and earnings, and reduced the likelihood of both poverty and incarceration in adulthood. The benefits of Head Start spending were larger when followed by access to better-funded public K12 schools, and the increases in K12 spending were more efficacious for poor children who were exposed to higher levels of Head Start spending during their preschool years. The findings suggest that early investments in the skills of disadvantaged children that are followed by sustained educational investments over time can effectively break the cycle of poverty.

# REDUCING INEQUALITY THROUGH DYNAMIC COMPLEMENTARITY: EVIDENCE FROM HEAD START AND PUBLIC SCHOOL SPENDING

RUCKER C. JOHNSON AND C. KIRABO JACKSON*

May 26, 2017

*We explore whether early childhood human-capital investments are complementary to those made later in life. Using the Panel Study of Income Dynamics, we compare the adult outcomes of cohorts who were differentially exposed to policy-induced changes in pre-school (Head Start) spending and school-finance-reform-induced changes in public K12 school spending during childhood, depending on place and year of birth. Difference-in-difference instrumental variables and sibling-difference estimates indicate that, for poor children, increases in Head Start spending and increases in public K12 spending each individually increased educational attainment and earnings, and reduced the likelihood of both poverty and incarceration in adulthood. The benefits of Head Start spending were larger when followed by access to better-funded public K12 schools, and the increases in K12 spending were more efficacious for poor children who were exposed to higher levels of Head Start spending during their preschool years. The findings suggest that early investments in the skills of disadvantaged children that are followed by sustained educational investments over time can effectively break the cycle of poverty.* (*JEL* I20, J20)

Children born to less-advantaged households and communities typically experience lower levels of educational attainment, employment, earnings, health, and well-being as adults than children born to more advantaged ones (Chetty, Hendren, and Katz, 2016). Differences between individuals from more- and less-advantaged backgrounds manifest early in childhood and tend to grow as children age (Fryer and Levitt, 2006; Currie and Thomas, 1999; McLeod and Kaiser, 2004; Heckman and Mosso, 2014). Accordingly, efficiently breaking the cycle of poverty may require early investments in the skills of disadvantaged children that are followed by sustained investments over time. We study whether early childhood investments for disadvantaged children that are followed up with increases in public school spending are particularly effective at improving their long-run outcomes.

Theory supports the prescription for early and sustained investments in human capital for

* Johnson: University of California, Berkeley and NBER; Jackson: Northwestern University and NBER. Johnson and Jackson are equal authors on this paper. We wish to thank the PSID staff for access to the confidential restricted-use PSID geocode data. We are grateful to Edward Zigler and seminar participants at the NBER education/children's meetings, the University of Michigan, UC-Berkeley, Penn, Brown, Vanderbilt University, University of Wisconsin/IRP Summer Workshop, and UC-Davis for helpful comments; and thank Martha Bailey & Andrew Goodman-Bacon, and Doug Miller & Jens Ludwig, for sharing data on 1960 county poverty rates and data from the National Archives & Records Administration. This research was supported by the National Institutes of Health and the Russell Sage Foundation (to Johnson), and the National Science Foundation (to Jackson).

disadvantaged children. Cunha and Heckman (2007) present a model of human capital production in which early human capital investments are complementary to those made later in life. If such complementarities exist, then early human capital investments make subsequent human capital investments more productive and may be ineffective if not followed by subsequent investments.[1] There is little experimental or quasi-experimental exploration into how the efficacy of policies that promote human capital in early life are affected by policies that promote it later in life and vice versa.[2] To fill this gap, we exploit two exogenous human capital investment "shocks" that occur at different points in the life course to explore whether early childhood human capital investments are complementary to those made later in childhood—that is, we explore whether human capital investments exhibit dynamic complementarity.

The first independent exogenous shock to human capital investment is the rollout of Head Start, which increased access to early childhood education and pediatric care for low-income children. The second independent exogenous shock to human capital investment is the implementation of court-ordered school finance reforms (SFRs) which (on average) increased the level of per-pupil spending at public K12 schools (Jackson, Johnson, and Persico, 2014). The first policy, Head Start, is the largest early childhood intervention program in the US. Head Start is a comprehensive, national, federally-funded early childhood program that was established in 1964 as part of Lyndon B. Johnson's "War on Poverty" to provide education, health care, nutrition, and other services to poor children before kindergarten entry. The second policy we exploit is court-ordered school finance reforms (SFRs). In the 1960s, the majority of public school spending was

[1] If early childhood investments for disadvantaged youth make subsequent investments in these children more productive, one can justify policies that redistribute resources toward disadvantaged children during their early years on *efficiency* grounds without any appeal to equity issues, fairness, or social justice (Heckman and Mosso, 2014).

[2] The most closely related work to ours are unpublished working papers that study the relationship between two plausibly exogenous human capital policies. Rossin-Slater and Wust (2016) examine whether the effects of access to pre-school differed among those who had access to home visits during infancy; Gilraine (2016) examines whether the benefits of accountability due to NCLB in later grades vary by exposure to accountability in earlier grades; and Malamud, Pop-Eleches and Urquiola (2016) examine whether the benefits of attending a better school vary by parental access to abortion near the time of conception. Another related group of unpublished papers examine whether the effect of educational interventions vary by measured ability. These papers interact a single policy shock with a potentially endogenous measure of ability or investment. These studies include Aizer and Cunha (2012) who explore whether the impacts of the launch of Head Start in 1966 had different effects for those with different measured stocks of early human capital. It also includes Garcia and Gallegos (2017) who examine whether the effects of randomized access to an early childhood education program varies among those with different levels of human capital (as measured before the intervention), and Lubotsky and Kaestner (2016) who explore whether individuals who start schools later benefit more from a year of schooling than those who start earlier.

funded through local property taxes, so that low-wealth areas tended to have lower per-pupil K12 spending levels than more affluent high-wealth areas. The court-ordered SFRs that began in the early 1970s (and continue to the present) changed the parameters of spending formulas. These changes reduced inequality in school spending and weakened the relationship between the level of public school spending and the wealth and income level of the district (Card and Payne, 2002; Murray, Evans, and Schwab, 1998; Hoxby, 2001; Jackson, Johnson, and Persico, 2014). Each of these two policies led to dramatic changes in the structure of public education in the United States. We explore the combined effects of the two.

To isolate the effects of these two major policies, we exploit temporal and geographic variation in exposure to these policy-induced investment "shocks" and analyze the life trajectories of individuals born between 1950 and 1976, and followed through 2013 using the Panel Study of Income Dynamics (PSID). Test scores have been the traditional focus of evaluations of Head Start and K12 spending. However, the effects of interventions on long-run outcomes may go undetected by test scores (e.g., Heckman, Pinto, and Savelyev, 2014; Deming, 2009; Jackson, 2012, forthcoming; Chetty et al., 2011; Ludwig and Miller, 2007). Consequently, we explore effects on an array of adult outcomes including educational attainment, earnings, poverty, and incarceration.[3]

We seek to identify the potential *interactive*, or synergistic, effects between early and later human capital investments. Identifying such interaction effects requires that we credibly identify the effects of each human capital investment individually. To identify the causal effect of early childhood investments, we exploit geographic variation in the timing of the rollout of Head Start across counties. In our preferred difference-in-difference models, we compare the adult outcomes of individuals who were from the same childhood county but were exposed to different levels of Head Start spending, because some were four years old when Head Start spending levels were low (or non-existent) while others were four years old when Head Start spending levels were higher. We present several empirical tests to show that the identifying variation in Head Start spending is unrelated to family, community, and other policy changes. We also show that our estimated Head Start effects are robust to (a) instrumental variable models that use only variation in Head Start spending due to the timing of Head Start rollout in the childhood county and (b) using within-

[3] In a recent related study, Carneiro and Ginja (2014) study the long-run effects of Head Start participation on health and behavioral problems. In our study we explore a wider array of adult outcomes such as earnings, family income, and child outcomes. Also, a key focus of our paper is on how the effectiveness of Head Start spending varies by the quality of the public schools students subsequently attend.

family, across-sibling variation in Head Start spending exposure.

To identify the causal effects of public K12 school spending, we exploit geographic variation in the timing of court-ordered SFRs. Following Jackson, Johnson, and Persico (2016), we predict the spending change that each district would experience after the passage of a court-mandated SFR based on the type of reform and the characteristics of the district before reforms. Using instrumental variables models, we examine whether SFR-exposed cohorts (young enough to have been in school during or after a SFR) have better outcomes relative to SFR-unexposed cohorts (those who were too old to be affected by a SFR) in districts predicted to experience larger reform-induced spending increases. We present several empirical tests showing that the within-district variation in per-pupil spending induced by SFRs is exogenous to other family, community, and policy changes in the district. We also show that our K12 school spending effects are robust to using within-family, across-sibling variation in SFR-induced K12 public-school spending.

To explore the relationship between early- and later-childhood human capital investments, we combine both identification strategies to estimate the effects of the interaction between Head Start spending and public K12 spending. We can test for dynamic complementarities based on two sources of variation. Namely, some districts experienced increases in school spending due to a SFR when Head Start was available in the county, while other districts experienced similar K12 spending increases when Head Start was not available. This fact allows one to test if the effects of K12 spending increases due to SFRs are higher with greater public pre-K investments than without them. Similarly, Head Start was rolled out in different counties both before and after the local school districts experienced increases in K12 spending due to SFRs. This fact allows one to test if the effects of Head Start spending are larger in areas that have higher levels of K12 spending due to the passage of a court-ordered SFR.

For children from low-income families, on average, increases in Head Start spending increased educational attainment and adult earnings and reduced the likelihood of both poverty and incarceration in adulthood. Consistent with Jackson, Johnson, and Persico (2016), increases in public school K12 spending improved this same array of outcomes in adulthood. We also find robust evidence of dynamic complementarity. Intent-to-treat estimates indicate that, for children from low-income families, on average, increasing Head Start spending by $1,000 per poor four-year-old (in the county) increases educational attainment by 0.096 years, increases adult wages by 1.9 percent, and reduces the likelihood of adult incarceration by 0.75 percentage points. However,

in districts at the 75th percentile of the K12 spending distribution, the same increase in Head Start spending increases educational attainment by 0.22 years, increases wages by 5.6 percent, and reduces the likelihood of incarceration by 2.2 percentage points (after accounting for the direct effect of higher levels of K12 spending). The dynamic complementarities are sufficiently large that the marginal effects of increases in Head Start spending are more than twice as large when K12 spending is at the 75th percentile than at the 25th percentile. The fact that the long-run benefits of Head Start spending depend on the subsequent level of K12 spending may help explain why some studies find positive effects of Head Start and others do not.[4] Looking at the marginal effects of K12 spending, for low-income children, increasing public K12 spending by 10 percent has small effects on educational attainment, adult wages, and incarceration when not preceded by Head Start. However, among low-income children exposed to Head Start spending at the 75th percentile of the distribution, the same 10 percent increase in K12 per-pupil spending increases educational attainment by 0.35 years, increases earnings by 13 percent, and reduces the likelihood of incarceration by 15 percentage points. The patterns of positive interaction effects between Head Start and K12 spending are robust across several models (including sibling comparisons) and are only present among children from low-income families. The effect of K12 spending was unrelated to the level of Head Start spending among non-poor children—for those children, increasing K12 spending by 10 percent increased years of education by 0.2 and earnings by 11.7 percent.

We find substantial long-run benefits of public early childhood investments and robust evidence of complementarities between early and later human capital investments for low-income children. The results imply, as suggested by Heckman and Mosso (2014), that investments in the early-childhood development of low-income children may not exhibit an equity-efficiency tradeoff, and that early and sustained investments in the skills of low-income children can be a cost-effective strategy to break the cycle of poverty.

The rest of the paper is organized as follows. Section II outlines our theoretical framework. Section III describes the Head Start program and court-ordered school finance reforms. Section IV presents the data used. Section V describes the empirical strategy. Section VI presents the results. Section VII presents conclusions and a summary discussion.

[4] For positive effects see Deming (2009), Ludwig and Miller (2007), Garces, Currie, and Thomas, (2002), Carneiro and Ginja (2014). For mixed effects see Zigler et al., (2011), Lipsey, Farran and Hofer (2015).

## II. Theoretical Framework

Research in developmental neuroscience highlights the importance of the preschool years in establishing the building blocks of subsequent human capital formation and the interconnectedness of cognitive, non-cognitive, and health formation (Shonkoff and Phillips, 2000). Evidence suggests that common developmental processes operate in the formation of cognitive, non-cognitive, and health capacities. Informed by this research and others, Cunha and Heckman (2007), theorize that skill development is an interactive, multistage process such that early-life human capital investments may cross-fertilize human capital investments made at later childhood stages, leading to developmental synergy effects. From this perspective, if the early childhood human capital investments provided by Head Start improve school readiness, they may facilitate better learning in the K12 system. If so, insofar as increased spending improves school quality, spending on Head Start and public K12 schools would be synergistic and would exhibit dynamic complementarities. We formalize this logic below.

Following the notation of Heckman (2007), we outline a model in which the technology of skills production is dynamic. Skills acquired when a child is *t* years old is [1] below

[1] $$\theta_{t+1} = f_t(h_t, \theta_t, I_t)$$

where *t*=1,2,...T, $\theta_t$ is a vector of skills at time *t*, parental capabilities are connoted by $h_t$, and investments during time *t* are connoted by $I_t$. Investments in time *t* ($I_t$) are construed broadly to include parental investments, schooling inputs (i.e., peers, teachers, etc.), and neighborhood and community inputs. For analytical convenience, $f_t$ is assumed to be strictly increasing in $I_t$. Dynamic complementarity in skill production arises when the stocks of capabilities acquired by period *t-1* ($\theta_t$) make investments in period $t$ ($I_t$) more productive, i.e.,

[2] $$(\partial\theta_{t+1})/(\partial\theta_t\partial I_t) > 0.$$

It is important to note that, unlike other studies, we do not seek to identify the skill production function parameter $(\partial\theta_{t+1})/(\partial\theta_t\partial I_t)$. Instead, we explore the closely-related ***policy*** question of how public human capital investments made in early childhood affect the efficacy of those made in later developmental stages of childhood. Consider that $\theta_t = f_{t-1}(h_{t-1}, \theta_{t-1}, I_{t-1})$. Because $\partial f_t/\partial I_t > 0$ from above, if [2] holds, then equation [3] below must also hold.

[3] $$(\partial\theta_{t+1})/(\partial I_{t-1}\partial I_t) > 0.$$

In words, dynamic complementarity in skill production implies that there is dynamic complementarity in human capital investments.[5]

In this paper, we test empirically for dynamic complementarity between early- and later-childhood human capital investments; in particular, between public investments in spending on early education for poor children (i.e., Head Start) and the public K12 system. We hypothesize that these two human capital policies may exhibit dynamic complementarity through a direct skill acquisition channel, and an indirect spillover channel.

The direct channel is what we call the "alignment" channel. This channel is predicated on the idea that the sequence of when skills are taught matters (Knudsen et al., 2006; Newport, 1990; Pinker, 1994) and the fact that K12 systems target students with a specific incoming skill level. Students above the target skill level may benefit less from the K12 system (the K12 system may spend valuable instructional time teaching skills they have already mastered), and students below this target incoming skill level may benefit less from the K12 system (the instruction may assume required skills that they do not possess). Given that poor children, on average, are less likely to be school-ready at kindergarten entry (Fryer and Levitt, 2004; Magnuson and Waldfogel 2005), Head Start spending, by increasing their skills, may bring them closer to the target such that they benefit more from subsequent investments experienced in the K12 education system. Furthermore, access to pediatric care (provided to Head Start participants) may promote this skill development (Levine and Schanzenbach, 2009; Cohodes, Grossman, Kleiner, Lovenheim, 2015).

Through this alignment channel, Head Start spending increases may not improve outcomes to the same degree in all contexts. In fact, in poorly-funded schools that may align instruction to a low-target skill level, Head Start participation could reduce alignment with the target level by increasing students' incoming skills above the target. In such a scenario, relative to their peers who did not attend preschool, any advantage in skill created by Head Start will diminish over time as children who attended Head Start receive redundant instruction, and their peers who lack access

[5] Note that, if early investments can increase the efficacy of later investments through mechanisms other than increasing skills, the converse does not *necessarily* hold. For example, suppose an increase in Head Start spending led to an increase in the supply of pre-K and K12 teachers (perhaps due to the creation of a new teacher-training program to meet the new demand). Then, increased Head Start spending would reduce the amount of money required to attract quality K12 teachers –thus making each marginal dollar in the K12 system more effective. Any such non-skills mechanism would generate spillover effects to non-Head Start participants that we can test for empirically. We present several empirical tests that such mechanisms are not a factor.

to preschool catch up in elementary school grades. That is, there may be fadeout and lower long-run Head Start effects for program participants who attend poorly-funded K12 schools. In sum, through this channel, on average, the effects of Head Start spending on poor children may be larger in well-funded K12 districts and could be negligible in poorly funded public school districts.

The indirect channel is through "spillover effects." In general, research has found that higher shares of low-performing peers or disruptive peers, and high levels of heterogeneity in ability levels in the classroom have deleterious impacts on student outcomes (see Sacerdote (2014) for an overview of this literature). By increasing the human capital of poor children, increases in Head Start spending may affect the subsequent peer composition of the K12 classrooms for *all* children in the county. If higher levels of Head Start spending reduce the likelihood of having low-achieving or disruptive peers or lessen the degree of heterogeneity in the classrooms, it could make it easier for the K12 school system to translate additional resources into improvements in outcomes. Neidell and Waldfogel (2010) provide evidence of this channel by documenting spillover effects from preschool between Head Start and non-Head Start children on math and reading achievement. Moreover, if teachers in the K12 system alter the alignment of their instruction toward an incoming higher-ability student (in light of a lower share of low-achieving students due to Head Start spending), the quality of K12 instruction could be affected for all students. Importantly, these spillover effects need not occur in the same classroom, because lower shares of students requiring remediation or special services may allow schools to allocate resources toward more productive inputs, which may benefit all students in the school.[6]

Through the hypothesized direct "alignment channel", both the potential direct effects of Head Start spending and dynamic complementarity will be experienced only by Head Start participants. However, through the indirect "spillover effects" channel, *all* children in K12 schools with former Head Start participants may experience the indirect dynamic complementarity effects, in addition to the direct effects experienced by Head Start participants.

[6] One of the program components of Head Start is teaching parenting skills; thus, another possible indirect channel is changes in parental quality. However, we argue that this potential mechanism is unlikely to be the dominant channel at work in this study because existing evidence provides much stronger support for the primary mechanisms operating through (a) enhancing literacy, numeracy, reasoning and problem-solving, and decision-making skills; (b) access to pediatric care; and (c) improved nutrition (Zigler, 2010; Currie and Neidell, 2007). We also present a test of this using within-family variation. In such tests, we find no indication that siblings of those exposed to higher-levels of Head Start spending have improved outcomes (Appendix I). This runs counter to the parental quality mechanism.

## III. BACKGROUND AND OVERVIEW OF HEAD START AND SCHOOL FINANCE REFORMS

We study the combined effects of two well-known, broad-reaching, publicly funded human capital interventions that were targeted largely to low-income children at different ages (Head Start and K12 public school spending). We present background on each in turn.

### III.A. *Background on Head Start*

Head Start was established in 1964 as part of Lyndon B. Johnson's "War on Poverty," and is a national, federally-funded, early-childhood program with the aim of improving the human capital of poor children. The Head Start curriculum aims to enhance literacy, numeracy, reasoning, problem-solving, and decision-making skills. Head Start includes educational efforts for both parents and children to enhance nutrition in the home and provides its own nutritious meals for the children. Participating children receive development screenings, and programs connect families with medical, dental, and mental health services.[7,8] Head Start also provides first-time parents with parenting strategies (Zigler et al., 2011). Head Start currently operates more than 19,200 centers and serves more than 900,000 children.[9] Current Head Start expenditures average about $8,700 per enrolled child (in 2015 dollars). This level of per-pupil spending is much lower than those at model preschool programs such as Perry Preschool or Abecedarian (Blau and Currie, 2006).[10] However, per-pupil Head Start spending levels are on the same order of magnitude as the average public K12 per-pupil spending, which is currently about $10,700 (in 2014 dollars).[11]

Because we seek to explore the effects of Head Start spending on longer-run adult outcomes (among those who are adults today), we study the effects of Head Start at the inception of the program (1965 through 1980). Head Start was initially launched as an eight-week summer-only program in 1965 and then became a primarily part-day, nine-month program in 1966. Head Start is mainly funded federally.[12] To open a new Head Start center, local organizations (typically

[7] http://www.acf.hhs.gov/programs/ohs/about/head-start

[8] An OEO report of 1967 documents Head Start accomplishments in the first two years on child health that include 98,000 eye defects treated; 900,000 cases of dental problems addressed (5 cavities per child); 740,000 without polio vaccinations received vaccines; and 1,000,000 were given measles vaccinations.

[9] See appendix Figure A1 for the national, annual enrollment in Head Start between 1965 and 2013.

[10] Head Start spending per enrollee is about 60 percent of spending levels observed in model preschool programs.

[11] There is considerable variability around this national average in individual states. States spending the least per-pupil included Utah ($6,555), Idaho ($6,791), Arizona ($7,208), Oklahoma ($7,672) and Mississippi ($8,130).

[12] Head Start funds were allocated to states proportionately based upon each state's relative number of children living in families with income below the poverty line and the relative number of public assistance recipients in each state. Head Start in collaboration with the Medicaid Early Pediatric Screening, Diagnosis, & Treatment Program (EPSDT) provided comprehensive prevention and treatment services to preschool children.

non-profit organizations, for-profit agencies, or school systems) apply to the federal government for grant funds. Grantees provide at least 20 percent of the funding. After approval, Head Start grants are awarded directly to applying organizations subject to three-year grant cycles.[13] Each grantee must comply with student-to-teacher ratio guidelines and other standards outlined in the Head Start Act. During the first 15 years of the program, the average student-to-teacher ratio in a Head Start classroom was roughly 17:1 (Zigler, 2010).[14] During this early era of the program, the majority of Head Start children were enrolled in part-day centers (as opposed to full-day programs, which are 6 or more hours per day such as Abecedarian), and often part-year (GAO report, 1981).[15]

To be eligible for Head Start participation children had to be four years old. At each center, at least 90% of enrollees had to be from families whose income was below the federal poverty line, and at least 10% of children had to have a disability.[16] Figure 1 plots the national Head Start enrollments as a percentage of the number of income-eligible four-year-olds between 1960 and 1994. This figure reveals key patterns that put our empirical work in perspective. First, the ratio of children enrolled in Head Start to the number of poor four-year-olds was as high as 90 percent in the very early years and then stabilized around 60 percent (current levels are estimated at 55 percent). This ratio is important to keep in mind as we interpret the magnitudes of our intent-to-treat estimates (presented in Section VI). It is also important to note that between 1965 and 1970, most of the enrollment in Head Start was in summer-only programs. However, from 1972 and after that, most enrollment was in full-year Head Start. This illustrates that the early rollout of Head Start represented *both* increases in Head Start participation *and* enhancements in the Head Start

[13] As documented in Ludwig and Miller (2007), the poorest 300 counties initially received grant assistance to apply for funding at the program's inception.

[14] This student to teacher ratio is higher than the prevailing student-to teacher ratios in the model preschool programs of the Perry Preschool (5.7 children per teacher), the Abecedarian Project (6 children per teacher), and Chicago Child Parent Center and Expansion Program (8-12 children per teacher) (Cunha, Heckman, Lochner and Masterov, 2006; Fuerst and Fuerst, 1993; Carneiro and Ginja, 2014). Note also the much smaller scale of these model programs as the Perry and Abecedarian programs each served just over 100 disadvantaged children.

[15] We are unable to identify which of these options a local Head Start center offered children who attended (part-day vs full-day; part-year vs. full-year). Summer-only programs were phased out by 1981 (Gibbs et al., 2011).

[16] Children who are 4 years old and live in poverty (i.e., family income below the federal poverty guidelines, or family is on public assistance programs AFDC or SSI) are eligible to be enrolled in the program; and beginning in 1972 (as part of the Economic Opportunity Act Amendment) at least 10 percent of children per center must have a disability (without any income cap on the eligibility of these children). In 1969, a provision was added allowing children from families above the poverty level to receive Head Start services for a fee. A fee schedule for non-poor participants in Head Start was required; fees were prohibited for families below the poverty line. The eligibility criteria was mostly unchanged during the period of the program we analyze (Source: 45 CFR (Code Federal Regulations), Parts 1301 to 1311, Early Childhood Learning and Knowledge Center: http://eclkc.ohs.acf.hhs.gov/hslc; www.eric.ed.gov; Zigler and Valentine, 1979).

programs themselves. Another notable pattern is the decline in Head Start enrollment between 1966 and 1970. During this period, full-year Head Start programs enrollment was increasing at the same time that summer-only program enrollment was declining (somewhat more rapidly).

Figure 1 also plots the share of 3- and 4-year-olds enrolled in full-time daycare over time (as reported in the Current Population Survey, 1960-1995). This figure highlights that Head Start rollout coincides with a period in which most children were not in formal, full-time pre-school, and also coincides with a general increase in the proportion of children aged 3 to 4 enrolled in full-time pre-school. In the context of the estimated effects of Head Start during this rollout period, the counterfactual option in the early years is primarily home care, as opposed to some other full-time pre-K program (as might be the case with present-day public pre-K expansions).

Because Head Start programs vary in quality, size, and scope, we use Head Start spending as a way to measure both the presence of the program and also the quality, size, and extent of the program. While Head Start spending per enrollee may seem like a natural proxy for quality, such a measure fails to capture changes in spending that work through expansions in access.[17] As such, because the target eligible population for Head Start is poor four-year-olds, our measure of Head Start spending is federal Head Start spending per *poor* four-year-old in the county. Between 1965 and 1980, the average county with a Head Start center spent about $4,000 per poor child and about $5,300 per enrollee (in year 2000 dollars). There is considerable variation in timing of the establishment of Head Start centers. However, in most counties, the *first* Head Start center was established between 1965 and 1970.[18] The geographic variation in the timing of the rollout of Head Start is central to our empirical strategy to isolate exogenous variation in Head Start spending across birth cohorts within a county.

[17] The expansion of Head Start involved both increases in the number of enrolled children and increases in spending per enrolled child. Head Start spending per enrollee increases do not capture increases in the total number of children affected by Head Start, so that spending per four-year-old in the county is a more appropriate measure. To illustrate this point, we collected data on Head Start spending per enrollee and Head Start spending per poor 4-year old at the state level between 2003 and 2014 (years for which both sets of data are available). Using within-state changes in spending over time, a 10 percent increase in spending per income-eligible four-year-old is associated with only a 0.243 percent increase in spending per enrollee, on average. However, it is also associated with a 1.2 percent increase in the number of Head Start participants, and a 6.5 percentage-point increase in the percentage of income-eligible poor four-year-olds enrolled (Table B2). While spending per income-eligible 4-year-old is sensitive to both increases in funding per enrollee and increases in total enrollment, increases in spending per enrollee are unrelated to increases in enrollment. These patterns make clear that for studying the rollout of Head Start, spending per enrollee would be an inappropriate measure, and that spending per income-eligible 4-year old is a much better measure.

[18] Figure A2 presents each county in the United States color-coded by the year of its first Head Start center.

### III.B. *Background on School Finance Reforms*

The other major human capital interventions we study are the increases in public K12 school spending caused by court-ordered school finance reforms (SFRs). In most states, before the 1970s, local property taxes accounted for most resources spent on K12 schooling (Howell and Miller, 1997). Because the local property tax base is typically higher in areas with higher home values, and there are high levels of residential segregation by socioeconomic status, heavy reliance on local financing contributed to affluent districts' ability to spend more per student. In response to large within-state differences in per-pupil spending across wealthy/high-income and poor districts, state supreme courts overturned school finance systems in 28 states between 1971 and 2010. Because of these court decisions, many states implemented legislative reforms that led to important changes in public education funding.[19] Most of these court-ordered SFRs changed the parameters of spending formulas to reduce inequality in school spending and weaken the relationship between per-pupil school spending and the wealth and income level of the district.

As pointed out in Hoxby (2001), the effect of a SFR on school spending depends on (a) the type of school funding formula introduced by the reform and, (b) how the funding formula introduced interacts with the specific characteristics of a district. To capture some of this complexity, we follow Jackson, Johnson, and Persico (2016) and categorize reforms into four types. Foundation plans guarantee a base level of per-pupil spending and are designed to increase per-pupil spending for the lowest-spending districts. Spending limit plans prohibit per-pupil spending levels above some predetermined amount. Such plans tend to reduce spending for high spending districts and may reduce long-run spending for all districts. Reward-for-effort plans match locally-raised funds for education with additional state funds (often with higher match rates for lower-income areas). Such plans tend to increase spending for all districts with larger increases in low-income districts. Equalization plans typically tax all districts and redistribute funds toward lower-wealth and lower-income districts. These reform/formula types are not mutually exclusive.

[19] The first of these successful cases is the California case, *Serrano v. Priest*, decided in 1971. Challenges to state school finance systems were argued on either equity or adequacy grounds. The early challenges (1971- mid 1980s) were won on equity grounds. For "equity cases," local financing was found to violate the responsibility of the state to provide a quality education to all children. "Equity cases" sought to weaken the relationship between the quality of educational services and the fiscal capacity of the district. The more recent challenges (late 1980s onwards) were mounted on adequacy grounds. "Adequacy cases" rely on the fact that most states have a constitutional provision requiring the state to provide some minimum "adequate" level of quality schools for all children (Lindseth, 2004) and were argued on the grounds that low per-pupil spending levels in certain districts meant that the state had failed to meet this obligation. Between 1970 and 1990, 10 and 4 states had court-ordered reforms argued on equity and adequacy grounds, respectively.

To illustrate how the introduction of different formula types affected districts by pre-reform income levels and pre-reform spending levels, Figures 2 and 3 present event-study plots of the natural log of per-pupil spending at the district level (after removing both district and year fixed effects). Year 0 is the first year of the first court order in the state, year "-5" is five years before the first court order, and year "5" is five years after the initial court order. For each court order, we link all formula changes that occurred within three years to that court-ordered SFR. Figure 2 shows the evolution of per-pupil spending for districts in the bottom and top quartiles of per-pupil spending in 1972 (the year preceding the first court-ordered SFR) after court orders that led to the implementation of different kinds of funding formula plans. Figure 3 presents similar plots for districts in the top and bottom quartiles of the state income distribution in 1963. Figures 2 and 3 show that court-ordered SFRs that lead to the implementation of different funding formulas had different effects on districts by pre-reform income and spending levels. That is, reforms that lead to "reward-for-effort" formulas tended to increase per-pupil K12 spending in all districts; spending limits had the most pronounced spending reductions in high-spending districts; foundation plans led to the largest spending increases in low-income districts; and equalization plans were more equalizing by pre-reform spending levels than by pre-reform income levels.

These systematic patterns allow us to predict how much K12 school spending will increase in each district as a function of the reform type introduced at the state level and the pre-reform characteristics of the district. Because these relationships are unrelated to the decisions made by individual districts or demographic shifts that may affect public school spending levels, we can use this prediction to isolate the causal relationship between reform-induced K12 spending increases and students' longer-run outcomes.

## IV. DATA

We compiled data on annual Head Start spending at the county level, and public K12 school spending at the school district level. The Head Start spending data come from the National Archives Record Administration, Inter-university Consortium for Political and Social Research, and Surveillance, Epidemiology, and End Results population data. These are combined to form a county-level panel of Head Start spending per poor 4-year-old in the county between 1965 and 1980. Public K12 education funding data come from several sources that are combined to form a panel of per-pupil spending for US school districts in 1967 and annually from 1970 through 2000

and linked to a database of SFRs from Jackson, Johnson, and Persico (2016).[20] To avoid confounding nominal with real changes in spending, we convert both Head Start and K12 school spending across all years to 2000 dollars using the Consumer Price Index (CPI).

Our individual-level data on long-run outcomes come from the Panel Study of Income Dynamics (PSID, 1968-2013), and our analysis sample includes individuals born between 1950 and 1976 who have been followed into adulthood. These PSID cohorts straddle both the rollout of Head Start programs across the country and the implementation of the early waves of court-ordered SFRs.[21] We include all information on PSID individuals between 1968 and 2013.[22] We linked persons in the PSID using their census blocks during childhood to school spending data, SFR data, and Head Start spending data.[23] We then match the earliest available childhood residential address to the school district boundaries that prevailed in 1969 to avoid complications arising from endogenously changing district boundaries over time. We outline the algorithm in Appendix C. We also merge in county-level characteristics from the 1960 Census, and information on the timing of other key policy changes during childhood (e.g., school desegregation, hospital desegregation, Title I, rollout of other "War on Poverty" initiatives and expansion of safety net programs—described in Section V) from multiple data sources.[24]

---

[20] The Census of Governments has been conducted every five years since 1972 and records school spending for every school district in the US. The Historical Database on Individual Government Finances (INDFIN) contains annual district finance data for a sub-sample of districts from 1967, and 1970 through 1991. After 1991, the Common Core data (CCD) School District Finance Survey (F-33) includes data on school spending for every school district in the US. Details on how these databases were compiled and the coverage of districts in these data are in Appendix B.

[21] The share of individuals potentially exposed to Head Start expenditures at age 4 increases significantly with birth year over the 1950-1976 birth cohorts analyzed in the PSID sample. Two-thirds of the sample grew up in a state that was subject to a court-mandated SFR between 1971 and 2000 (the first court order was in 1971).

[22] The PSID maintains high wave-to-wave response rates of 95-98%. Studies have concluded that the PSID sample of heads and wives remains representative of the national sample of adults (Fitzgerald, Gottschalk, Moffitt , 1998a,b; Becketti et al, 1988). Additionally, we perform a supplementary analysis of sample attrition in the PSID, and find no evidence of selective attrition among our study sample (Appendix Table G5). In particular, among original sample children, baseline 1968 family & county characteristics do not jointly significantly predict the likelihood of attrition or the likelihood of being observed as an adult.

[23] The PSID began interviewing a national probability sample of families in 1968. These families were re-interviewed each year through 1997, when interviewing became biennial. All persons in PSID families in 1968 have the PSID "gene," which means that they are followed in subsequent waves. When children with the "gene" become adults and leave their parents' homes, they become their own PSID "family unit" and are interviewed in each wave. The original geographic cluster design of the PSID enables comparisons in adulthood of childhood neighbors who have been followed over the life course. Moreover, the genealogical design implies that the PSID sample today includes numerous adult sibling groupings who have been members of PSID-interviewed families for more than four decades. We include both the Survey Research Center component and the Survey of Economic Opportunity component, commonly known as the "poverty sample," of the PSID sample.

[24] The data we use include measures from 1968-1988 Office of Civil Rights (OCR) data; 1960, 1970, 1980, and 1990 Census data; 1962-1999 Census of Governments (COG) data; Common Core data (CCD) compiled by the National Center for Education Statistics; Regional Economic Information System (REIS) data; county-level Title I/ESEA

We define low-income children as those whose average parental income (between ages 12 and 17) fell in the bottom quartile.[25] Among cohorts born between 1963-1976 for whom parental income at age four is observed, roughly 80 percent of those whom we classify as low-income were below the federal poverty line at age four, and 93 percent of those who were below the poverty threshold at age four are classified as low-income by our definition. The analytic sample includes 13,381 individuals from 4,684 childhood families, 1,431 school districts, 1,070 counties, across all 50 states. From this point forward, we refer to children who are low income as "poor" children, and those not from low-income families (as defined above) as "non-poor" children. We examine a broad range of adult outcomes. These include 1) educational outcomes—whether graduated from high school, years of completed education; 2) labor market and economic status outcomes (in real 2000 dollars)— log wages, family income, annual incidence of poverty in adulthood[26] (ages 20-50); and 3) criminal involvement and incarceration outcomes—whether ever incarcerated (jail or prison) and the annual incidence of incarceration in adulthood. Table 1 contains descriptive statistics for various childhood measures and adult outcomes in our analytic sample.

## V. EMPIRICAL STRATEGY

We aim to uncover the causal effects of spending on Head Start programs, public school K12 spending, and the effects of the interaction between the two. To this aim, we exploit policy-induced changes in Head Start and public K12 education spending that are unrelated to child family- and neighborhood-level determinants of adult outcomes. Due to the complexities of the causal effect of each kind of education spending (early childhood versus K12), we describe each source of variation in turn, and then later discuss how we combine the two in our empirical models.

### V.A. *Identifying the effects of Head Start Spending*

Our measure of Head Start spending is total federal Head Start spending in a county per poor four-year-old (in 2000 CPI-adjusted real dollars). Our research design takes advantage of the

spending (NARA); the comprehensive case inventory of court litigation regarding school desegregation over the 1955-1990 period (American Communities Project), and major plan implementation dates in large districts (compiled by Welch/Light); and American Hospital Association's Annual Survey of Hospitals (1946-1990) and the Centers for Medicare Provider of Service data files (dating back to 1960s) to identify the precise date in which a Medicare-certified hospital was established in each county of the US (an accurate marker for hospital desegregation compliance).

[25] Because the earliest year in which parental income is available is 1967 due to when the PSID data collection started, we cannot observe family income at age four for those born before 1963. However, we can observe average family income during adolescence (ages 12 through 17) for all individuals in our analytic sample, which serves as a good permanent income measure. We use this to form our group of likely Head Start eligible individuals.

[26] Based on the family income-to-needs ratio and federal poverty thresholds by family structure and household size.

staggered introduction across geographic areas of Head Start programs and the resulting spending increases during the program's rollout. Before the rollout of Head Start to an area, there is no spending on Head Start. However, after the introduction of Head Start in a county, spending levels typically increase for several successive years. Figure 4 shows an event-study plot of Head Start spending per poor-four-year-old before and after rollout in areas that had high and low Head Start spending in 1980 (the end of the sample period under study). Note that year "zero" is the year of the establishment of the first Head Start center in a county.

In the high-spending counties, once the first center is established, spending per poor four-year-old increases rapidly. As expected, the increase is much larger in the high-spending counties (from zero to about $5,000 per poor 4-year old) than the low-spending counties. However, spending is highest during the first year and then falls after that. This initial increase and subsequent fall is an artifact of the large national enrollment in summer-only programs that were phased out in the following years (Figure 1). The initial increase in Head Start spending due to the summer-only programs is also evident in counties with low Head Start spending in 1980. In essence, almost all counties experienced a transitory increase in Head Start spending, due to the ubiquitous introduction of summer-only programs that falls over time. However, high-spending counties expanded enrollment (and spending) in full-year programs that was sustained over time, while the low-spending counties did not increase spending on full-year programs and reverted to near zero Head Start spending within four years.

The left panel of Figure 4 reveals that, among children born within a few years of each other in the same county, some were four years old when there was no Head Start spending in their county, and others were four years old at the end of the phase-in stage when spending levels were high. If higher levels of Head Start spending improve outcomes, one should observe that (a) the post-rollout cohorts should have better outcomes than the pre-rollout cohorts, and (b) improvements between pre- and post-rollout cohorts should be larger in counties with larger sustained increases in Head Start spending. The right panel of Figure 4 reveals exactly this pattern for years of educational attainment (measured in adulthood) among poor children. The event study shows that areas with small increases (dashed grey line) and those with large increases in Head Start spending (solid black line) were on the same trajectory among cohorts who were older than four years old when the first Head Start center was established (i.e., years -8 through year 0). However, the post-rollout cohorts have much better outcomes in high Head Start spending counties

than in low-spending counties. This provides a graphical representation of our empirical strategy.

Our preferred difference-in-difference (DiD) strategy uses all this variation in timing and dosage. That is, we compare the differences in long-run outcomes across birth cohorts from the same childhood county that experienced larger increases in Head Start spending at age 4, to the differences in outcomes across the same birth cohorts within other childhood counties that experienced small (or no) increases in Head Start spending at age 4. These DiD type comparisons are implemented in a regression framework by estimating [4] by Ordinary Least Squares (OLS).

[4] $$Y_{icb} = \beta \cdot HS_{cb}^{age\,4} + \gamma \cdot C_{icb} + \theta_c + \tau_b + \varepsilon_{icb}.$$

In [4], $Y_{icb}$ is the outcome of individual *i*, from childhood county *c*, in birth cohort *b*. The variable of interest ($HS_{cb}^{age\,4}$) is Head Start spending per poor four-year-old in county *c* (in year 2000 dollars), when birth cohort *b* was age 4. To rely only on within-county variation in Head Start spending across cohorts, [4] includes childhood county fixed effects ($\theta_c$), and to account for cohort effects we include birth-year fixed effects ($\tau_b$). We also include an extensive set of childhood-family and individual characteristics, and county-level coincident policy changes as control variables ($C_{icb}$) that we detail in Section V.C. The idiosyncratic error term is $\varepsilon_{icb}$.

There are two identifying assumptions. First, counties that experienced increased Head Start spending over time (where most of the variation occurs at rollout) were not already on a trajectory of improving or deteriorating outcomes over time. Second, counties that saw larger or smaller increases in Head Start spending did not also undergo other unobserved changes that would also affect outcomes. Figure 4 suggests that the first condition is satisfied. Furthermore, in Section VI.B we present evidence that support the validity of these identifying assumptions.

### V.B. *Identifying the effects of K12 School Spending*

Our measure of K12 public school spending during childhood, $ppe_{idb}^{5-17}$, is the *natural log* of average public K12 school spending per-pupil (in real 2000 dollars) during school-age years (ages 5 through 17) in an individual's childhood school district.[27] We refer to this as K12 spending. Individuals who turned 17 years-old during the year of the passage of a court-ordered SFR in their state should have completed secondary school by the time reforms were enacted. Such cohorts

[27] The average level of district per-pupil spending across all school-age years provides a summary measure of the level of financial resources available in the individual's childhood school district during all their school-going years (ages 5 through 17 corresponding to expected grades K12). We use the natural log of this average measure to capture the fact that school spending likely exhibits diminishing marginal product.

(and older cohorts) are "SFR *unexposed*". Individuals who turned 16 years old or younger during the year of the passage of the first court-ordered SFR in their state would likely have attended primary or secondary school when reforms were implemented. Such cohorts are "SFR *exposed.*" One can estimate the SFR exposure effect on outcomes for individuals from a particular district by comparing the change in outcomes between SFR-exposed and SFR-unexposed birth cohorts from that district. Some districts experienced larger spending increases due to a court-ordered SFR than others did. Accordingly, we test for a causal effect of per-pupil spending during childhood by testing whether the difference in outcomes between SFR-exposed and SFR-unexposed cohorts from the same school district (i.e., the SFR exposure effect) tends to be larger for those districts that experienced larger reform-induced K12 spending increases (i.e., a SFR dose-response effect). Our identifying assumption is that the spending changes caused by the reforms within districts were unrelated to other district-level changes that could have affected adult outcomes directly.

Following Jackson, Johnson, and Persico (2016), we quantify the relationship between K12 spending and adult outcomes by using only the variation above in school spending associated with the passage of a court-mandated SFR. Specifically, using the PSID, we estimate equation [5] by 2SLS. All common variables are defined as in [1].

[5] $$Y_{idcb} = \beta \cdot \widehat{ppe}_{idb}^{5-17} + \gamma \cdot C_{idcb} + \theta_d + \tau_b + \varepsilon_{idcb}.$$

To rely only on variation across birth cohorts within districts, we include school district fixed effects ($\theta_d$); to account for time trends and cohort effects, we include birth-year fixed effects ($\tau_b$); and to account for life cycle effects, we include flexible controls for age (cubic). Our endogenous regressor is $ppe_{idb}^{5-17}$, and $\widehat{ppe}_{idb}^{5-17}$ are fitted values from a first stage.

The excluded instruments in the first stage are measures of exposure to a SFR interacted with measures of dosage (to account for the fact that some districts have larger reform-induced spending increases than others). Our exposure measure, $SFRExp_{icb}$, is the number of years individual *i* in birth cohort *c* from childhood district *d* is expected to have been in school after the passage of the first court-ordered SFR in their home state. This exposure measure varies at the state birth-cohort level and goes from 0 (for those who were age 17 or older the year of the state's first court ordered SFR) to 12 (for those who were ages 5 and younger the year of the state's court ordered SFR). To capture variation in dosage conditional on exposure, in the first stage we also include the two-way interaction between $SFRExp_{idb}$ and a district-level predictor of the spending change caused by the state court-ordered SFR in that district ($\widehat{dose}_d$). More formally, the first

stage regression is as in [6] below

[6] $$\widehat{ppe}_{idb}^{5-17} = \pi_1\left(SFRExp_{idb} \times \widehat{dose}_d\right) + \pi_2(SFRExp_{idb}) + \gamma_1 \cdot C_{idcb} + \theta_{d1} + \tau_{b,1}.$$

It is important that we use an exogenous predictor of $\widehat{dose}_d$ that is unrelated to the potentially endogenous decisions made by districts after reforms. Our measure, $\widehat{dose}_d$, is a weighted average of reform type, pre-reform district income levels, pre-reform district spending levels and their interactions. By construction, $\widehat{dose}_d$ is unrelated to endogenous decisions made by districts after reforms. To form $\widehat{dose}_d$, we use the full universe of school districts and regress per-pupil spending on (a) indicators for years of SFR exposure, interacted with reform type, interacted with pre-reform spending levels in 1972; and (b) indicators for years of SFR exposure, interacted with reform type, interacted with pre-reform median income levels in 1963, and region-specific year fixed effects. We then take the fitted values from this regression to obtain a predicted reform-induced spending change for each district (based on reform type implemented by the state, district spending prior to reforms, and district income levels prior to reforms). See Appendix E for details. Because $\widehat{dose}_d$ is estimated using *all* school districts while we estimate effects using districts represented in the PSID sample, our approach is a two-sample-2SLS.[28]

To show that $\widehat{dose}_d$ captures meaningful variation in K12 spending caused by court-mandated SFRs, Figure 5 shows the evolution of K12 spending among individuals in the PSID sample from districts with high predicted dosage (i.e. $\widehat{dose}_d$>0 ) and those with no predicted increases (i.e. $\widehat{dose}_d \leq 0$).[29] We create "event-time" indicator variables denoting the year an individual turned 17 minus the year of the first court order in the childhood state of individual *i*. The "-5" cohort are individuals who were 22 years old at the passage of a court-ordered SFR, the "-1" cohort was 18 years old at the passage of a court-ordered SFR, and the "5" cohort was 12 years old at the passage of a court-ordered SFR in their state. We then estimate a regression model predicting school-age K12 spending as a function of year fixed effects, district fixed effects, and the event-time indicators interacted with whether the district is predicted to have increased K12

[28] This approach was popularized by Angrist and Krueger (1992) and has been used in several other settings (e.g., Bjorklund and Jantti, 1997; Currie and Yelowitz, 2000; Dee and Evans, 2003).

[29] Roughly two-thirds of districts in reform states are predicted to experience spending increases in the first 8 years due to court-ordered SFRs. As one can see from Figure 5, because K12 spending tended to increase in states following court-ordered SFRs in general, there are small increases in K12 spending within 12 years post reform even in districts with predicted initial decreases. As such, we refer to all districts as having high- or low-predicted increases.

spending due to the passage of a court-ordered SFR. Because the outcome is in logs, the values represent percent changes in average school-age spending relative to the cohort from the same district that was 17 the year of the first court-ordered SFR.

Consistent with the timing of court-ordered SFRs being exogenous to underlying trends in school spending, both districts with lower and higher predicted dosage were on similar pre-reform trajectories as similar districts in non-reform states. Consistent with $\widehat{dose}_d$ isolating real variation in dosage, cohorts that turned 5 years old during the year of the initial court order (cohort 12) in districts with $\widehat{dose}_d > 0$ experience a 19 percent increase in school-age per-pupil spending, while the same cohorts in districts with $\widehat{dose}_d \leq 0$ experience a 5 percent increase. The timing of the initial court-ordered SFR in the state interacted with the predicted reform-induced spending increase for the district (based on state reform type interacted with pre-reform district characteristics) likely isolates exogenous variation in school spending. We present additional tests to support the validity of this approach in Section VI.B.

If our identification strategy is valid and K12 spending affects outcomes, outcome differences across exposed and unexposed cohorts should follow similar patterns to those of K12 spending. The right panel of Figure 5 shows this for years of educational attainment. Areas that had small (gray line) and large (black line) reform-induced increases in K12 spending were on similar trajectories among the unexposed cohorts (years -8 through year 0). However, the post-SFR cohorts (years 0 through 12) experienced much larger increases in years of education in the high-predicted K12 spending increase districts than in the low-predicted K12 spending increase districts. This figure depicts graphically the variation that undergirds our identification strategy.

### V.C. *Testing for Dynamic Complementarity*

To test whether the marginal effect of increased Head Start spending varies by the level of K12 spending and vice versa, we estimate the effects of public pre-K and K12 spending on adult outcomes with the inclusion of the interaction between Head Start spending at age 4 ($HS_{icb}^{age\,4}$) and the natural log of public K12 spending between the ages of 5 and 17 ($ppe_{idb}^{5-17}$). All models are estimated separately for poor and non-poor children, as we do not expect to find significant effects of Head Start spending nor evidence of dynamic complementarity among non-poor children (at least through direct channels as they are not income-eligible for Head Start). We define $\text{INT}_{idb} = \left(HS_{icb}^{age\,4} \times ppe_{idb}^{5-17}\right)$. We use the DiD variation (i.e., the within-county variation) in Head Start

spending ($HS_{icb}^{age\ 4}$), and instrument for both K12 spending ($ppe_{idb}^{5-17}$) and the interaction between Head Start and K12 spending ($INT_{idb}$). Our excluded instruments are all the two-way and three-way interactions between (a) the number of school-age years of exposure to a court-ordered SFR, (b) predicted dosage, and (c) Head Start spending per four-year-old when the individual was age 4.[30] Because a school district may be a smaller unit of observation than a county, all models include district fixed effects (which subsumes the childhood county fixed effects). The resulting model is as in [7], where $\widehat{ppe}_{idb}^{5-17}$ and $\widehat{INT}_{idb}$ are fitted values from first-stage regressions.

[7] $$Y_{icb} = \beta_{HS} \cdot HS_{cb}^{age\ 4} + \beta_{k12} \cdot \widehat{ppe}_{idb}^{5-17} + \beta_{int} \cdot (\widehat{INT}_{idb}) + \gamma \cdot C_{icb} + \theta_d + \tau_{bdi} + \varepsilon_{idb}.$$

The interaction effect between pre-K and K12 spending can be identified in [7] because (a) among counties that faced similar increases in Head Start spending, some were located in school districts that experienced larger (or smaller) increases in K12 spending due to the passage of a court-ordered reform, and (b) among cohorts from districts that faced similar increases in K12 spending due to the passage of a court-ordered reform, some grew up in counties that had higher (or lower) levels of Head Start spending when those cohorts were age 4.

To further reduce the possibility of confounding effects, vector $C_{idb}$ includes a variety of individual, childhood family, and childhood county controls. These include parental education and occupational status, parental income, mother's marital status at birth, birth weight, child health insurance coverage, gender, and the adult economic and incarceration outcomes include flexible controls for age (cubic). $C_{idb}$ also includes birth-year fixed effects by region and race, birth-cohort linear trends interacted with various 1960 characteristics of the childhood county (poverty rate, percent black, average education, percent urban, and population size). Also, to avoid confounding our effects with that of other policies that overlap our study period, $C_{idb}$ includes controls for county-by-year measures of school desegregation, hospital desegregation, community health centers, state funding for kindergarten, Title I school funding, imposition of tax limit policies, average childhood spending on food stamps, Aid to Families with Dependent Children, Medicaid, and unemployment insurance (Johnson, 2013; Chay, Guryan, & Mazumder, 2009; Hoynes, Schanzenbach, and Almond, 2016). Standard errors are clustered at the state level.

[30] Specifically, $\widehat{ppe}_{idb}^{5-17} = \pi_{11}(SFRExp_{idb} \times dose_c) + \pi_{12}(SFRExp_{idb}) + \pi_{13}(SFRExp_{idb} \times dose_c) \cdot HS_{cb}^{age\ 4} + \pi_{14}(SFRExp_{idb}) \cdot HS_{cb}^{age\ 4} + \pi_{14}(dose_c) \cdot HS_{cb}^{age\ 4} + C_{idb} + \theta_{d1} + \tau_{b1}$.
and $\widehat{INT}_{idb} = \pi_{21}(SFRExp_{idb} \times dose_c) + \pi_{22}(SFRExp_{idb}) + \pi_{23}(SFRExp_{idb} \times dose_c) \cdot HS_{cb}^{age\ 4} + \pi_{24}(SFRExp_{idb}) \cdot HS_{cb}^{age\ 4} + \pi_{24}(dose_c) \cdot HS_{cb}^{age\ 4} + C_{idb} + \theta_{d2} + \tau_{b2}$.

To provide visual evidence of complementarity effects, Figure 6 plots the estimated changes in years of educational attainment for cohorts before and after a court-ordered SFR for districts with high predicted spending increases (i.e., $\widehat{dose}_d$>0 ) and those with no predicted increases (i.e., $\widehat{dose}_d \leq 0$), separately for children with and without a local Head Start center at age 4. The left panel shows that SFR-treated cohorts and SFR-untreated cohorts experienced similarly small changes in educational attainment in districts that had small increases in K12 spending and were not exposed to Head Start at age four (grey line). However, among cohorts that had county Head Start spending at age four, school-age years of exposure to SFRs led to increases in educational attainment relative to those who were not exposed to SFRs. This pattern is consistent with Head Start making even small increases in K12 spending effective for low-income children. However, if the two policies are complementary, one should see similar patterns and greater increases in completed education for large increases in K12 spending. This is precisely what we document in the right panel of Figure 6. Here we see that in districts that experienced large increases in K12 spending after a SFR, exposed cohorts achieve more years of education than unexposed cohorts, and there is a dose-response relationship with the number of school-age years of exposure to larger reform-induced spending increases. Importantly, the relative increase in years of education is larger among those SFR-exposed cohorts that were from counties with a Head Start center at age four than among those SFR-exposed cohorts that did not have a Head Start center at age four. Furthermore, if one compares the effects across the two panels, one can see that the benefits of Head Start spending (the difference between the grey and black line in each panel) are larger among exposed cohorts that experience larger K12 spending increases. In sum, Figure 6 presents flexible semi-parametric evidence that Head Start and K12 school spending exhibit dynamic complementarity. The lack of any differential pre-trending in either panel illustrates that the parallel trends assumption likely holds not just for each policy (as previously shown in Figure 4 and 5) but also for the *interaction* between the two policies.

## VI. RESULTS

We present results from specification [7] that exploit all the within-county across-cohort variation in Head Start spending and instruments for K12 public school spending at the district level using the SFR instruments. To facilitate interpretation of the base effects of K12 spending and Head Start spending when the interaction between the two is included, both K12 spending and

Head Start spending are centered on their respective means. Thus, the coefficient on Head Start is the marginal effect of Head Start spending at the average level of K12 spending, and the coefficient on K12 spending is the marginal effect of K12 spending at the average level of Head Start spending. To organize our discussion, we first discuss the base effects of K12 spending (in logs) and Head Start spending, present empirical evidence that these estimated base effects are unbiased, and then discuss the estimated interaction effects. We present our estimated effects on education outcomes, followed by adult economic outcomes, and finally incarceration.

### VI.A. *Estimating the Base Effects of Head Start and K12 Spending*

Table 2 presents the estimates from [7] for low-income (bottom income quartile during childhood) children. Column 1 presents the effects on the probability of graduating from high school. The coefficient on Head Start spending per poor four-year-old is 0.01984 ($p$-value<0.01). This indicates that increasing Head Start spending per poor 4-year-old in the county by $1,000 (roughly a 25 percent increase) increases the likelihood of graduating from high school by 1.9 percentage points for a poor child exposed to the average level of K12 spending. Given that the average level of Head Start spending, conditional on having any Head Start program in the county, is about $4,000, this implies that, for poor children, having access to the average Head Start program increased the likelihood of graduating from high school by roughly 7.6 percentage points.

Increases in Head Start spending can affect outcomes through increasing Head Start participation, increases in the quality and scope of Head Start services, and can also indirectly affect outcomes through peer effects in the K12 system due to having better-prepared schoolmates. While existing studies have focused on the effect of *enrolling in* Head Start as participants, we estimate the effect of Head Start spending on *all* eligible children. Because there are multiple channels through which spending effects may emerge, we provide a sense of how our spending effects relate to the participation effects in the extant literature. Unfortunately, the PSID survey did not collect information regarding Head Start participation during the years in question.[31] However, we can approximate this margin using national data. Between 1965 and 1980, Head Start enrollment (full-time or part-year) accounted for roughly 66 percent of all eligible four-year-olds

[31] The PSID survey data employed in Garces, Currie, and Thomas (2002) are retrospective data collected in the 1995 wave for individuals born after 1965 (i.e., who were 4 years old after 1969). Thus, Head Start participation information was not collected retrospectively for the cohorts who were age four during the ramp-up period in which most of our variation is derived. See Appendix F for further discussion. The implied participation effects using these data are similar to those we assume here.

(Figure 1). Because centers can enroll 10 percent of non-poor children and must include some disabled children, the participation rate among income-eligible children could have been as low as 60 percent. Roughly 80 percent of poor children born after 1962 in the PSID resided in a county with a Head Start center at age four during this period (this is consistent with national figures). Assuming that only children with a Head Start center in their local area at age four will participate, this implies a Head Start participation rate of about 0.6/0.8=0.75, conditional on having a Head Start center in the county. We take this as our "ballpark" estimate of the increase in the likelihood of Head Start participation (among low-income children) due to the rollout of the average Head Start center in the county during our study period.

If all of our estimated intention-to-treat effect of having access to a Head Start center was due to Head Start enrollment (and there were no spillover effects to other low-income children), our assumed participation margin effect implies a treatment-on-the-treated effect of 0.076/.75=0.105, or 10.5 percentage points. This is similar to the estimated enrollment effect of Head Start in existing studies.[32] However, most existing studies of Head Start focus on full-year Head Start programs. If one considers the effect of only full-year Head Start, the average enrollment rate among eligible children was about 40 percent between 1965 and 1980. This implies a full-year Head Start participation rate of about 0.4/0.8=0.5 conditional on having a Head Start center in the county at age 4. If all of our estimated intention-to-treat effect was due to *full-year* Head Start enrollment, this assumed full-year Head Start participation margin effect implies a treatment-on-the-treated effect on the likelihood of high school graduation of 0.076/.5=0.152, or a 15.2 percentage-point increase. This is in line with the larger of the participation margin effects in the literature, or with the smaller participation effects in the literature if there are modest spillovers. In sum, our spending effects can be explained entirely by an enrollment effect using the range of estimates in the existing literature. However, we cannot rule out that some modest portion of our effects are driven by (a) improvements in the quality and scope of Head Start centers (full day versus half day, full time versus summer only, better teachers, etc.), and (b) spillovers from Head Start participants to low-income non-participants in the K12 school system.

[32] For example, Currie et al (1995) find that participating in Head Start increases the high school graduation rates for white by 20 percentage points, with no statistically significant effect for blacks. Deming (2009) finds that Head Start participation increases high school graduation by 11 percentage points for blacks with a small effect for whites, and increases high school graduation by 16 percentage points for those with low maternal test scores. Weikart, Marcus and Xie (2000) find that the average effect is 14 percentage points.

As expected, the coefficient estimates for K12 spending are very similar to those presented in Jackson, Johnson, and Persico (2016). The coefficient on the log of K12 spending during the school age years is 0.5956 ($p$-value<0.01). That is, increasing K12 school spending (across all 12 school-age years) by 10 percent increases the likelihood of high school graduation by 5.9 percentage points for a poor child exposed to the average level of Head Start spending. Relative to baseline, this is an 8.3 percent increase. The estimates indicate that increasing Head Start spending by $3,000 would have roughly the same effect on high school graduation as increasing K12 spending by 10 percent across all school-age years (for poor children).[33]

Column 2 presents a similar pattern for completed years of education for low-income children. Increasing Head Start spending per poor 4-year old in the county by $1,000 increases the years of educational attainment by 0.096 years for a poor child exposed to the average level of K12 spending. As such, at average Head Start spending levels of about $4,000, a Head Start center is estimated to increase years of education by roughly 0.38 (just over a third of a year). Increasing school-age K12 spending by 10 percent increases the number of years of completed education by 0.237 years for a low-income child exposed to the average level of Head Start spending.

Results for non-poor children (top 3 income quartiles during childhood) are in Table 3. The estimated K12 spending effects on the education outcomes are positive and sizable, but not statistically significant. The point estimates indicate that increasing K12 spending in the district by 10 percent increases the likelihood of high school graduation by 3.2 percentage points, and increases years of educational attainment by about 0.2 years for a non-poor child exposed to the average level of Head Start spending. These estimated effects are similar in magnitude to the effects for poor children, but are less precise. Accordingly, while one cannot reject that the marginal effects of K12 spending on the education outcomes are different for low-income and non-poor children, one can also not reject that the estimated effects are zero for non-poor children.

While the estimated effects of K12 spending are similar across the two groups, the estimated effects of Head Start are very different (as expected). In contrast to the large effects for low-income children, increasing county Head Start spending has very small, insignificant effects on non-poor children exposed to average levels of K12 spending. For both education outcomes, one *can* reject that the marginal effect of Head Start is the same for poor and non-poor children;

[33] During the sample period, a 10 percent increase in K12 spending is roughly equal to increasing per-pupil K12 spending by $480 each year over 12 years (about $4300 in present value terms assuming a 7 percent interest rate).

one cannot reject that the effect on the non-poor is zero; and one can reject that the effect for poor children is zero. This suggests that (a) there are no spillover effects of Head Start spending on non-poor children, and that (b) increases in Head Start spending are not associated with other policies that improve the outcomes of non-poor children. This bolsters the credibility of the research design. Note that the small, insignificant effects found for non-poor children does not rule out spillover effects on poor children who did not attend Head Start.

The adult economic outcomes we examine are wages, annual family income, and the annual incidence of poverty between the ages of 20 and 50. Our models that analyze adult economic outcomes (such as wages and annual family income) use all available person-year observations for ages 20–50 and control for a cubic in age to avoid confounding life cycle and birth cohort effects. Columns 3, 4, and 5 in Table 2 present these results for children from low-income families. Looking at wages, the coefficient on the log of public K12 school spending is 0.927 ($p$-value<0.1) and that on Head Start spending per poor 4-year-old is 0.0193 ($p$-value<0.01). That is, for children from low-income families exposed to average levels of Head Start spending, increasing K12 spending by 10 percent is associated with about 9.3 percent higher adult wages. Similarly, for these same children, increasing Head Start spending by $1,000 per poor 4-year-old is associated with 1.93 percent higher wages in adulthood. At the average level of Head Start spending following the program's rollout, this implies an average Head Start rollout effect of 7.75 percent higher wages due to the expansion of public pre-K availability for poor children.

Column 3 of Table 3 presents the effects on adult wages for non-poor children. Similar to the education outcomes, there are positive effects of K12 spending, but no effect of Head Start spending on the wages of adults who were from non-poor families. The coefficient on the log of K12 public school spending is 1.173 ($p$-value<0.05), and that on Head Start spending per poor 4-year-old is 0.00617 ($p$-value>0.1). That is, for children from non-poor families exposed to average levels of Head Start spending, increasing K12 spending by 10 percent is associated with 11.7 percent higher earnings between the ages of 20 and 50. However, consistent with no Head Start spillover effects on non-poor children, increasing Head Start spending is associated with no difference in earnings between the ages of 20 and 50.

The pattern of estimates for family income and annual incidence of poverty in adulthood in columns 4 and 5 of Tables 2 and 3 mirror those for adult wages. Head Start spending is associated with large, statistically significant improvements in the adult economic outcomes of

poor children (Table 2), and has small, insignificant effects on the adult outcomes of non-poor children (Table 3). However, increases in public K12 spending are associated with sizable improvements in the economic outcomes of all children, on average. The one exception to this pattern is adult poverty, for which neither Head Start spending nor K12 spending has an effect among non-poor children (likely because baseline rates are low for this population). Indeed, the likelihood of being in poverty at age 30 is only 5 percent for non-poor children compared to 18 percent for poor children.

The final outcome we examine is the probability that an individual has ever been incarcerated (Column 6 of tables 2 and 3). As with adult poverty, we find no effect of either Head Start or K12 spending on the likelihood of adult incarceration among non-poor children. We, therefore, focus our discussion on the outcomes of poor children. For these children, a $1,000 increase in Head Start spending per poor four-year-old reduces the likelihood of being incarcerated by 0.7 percentage points ($p$-value<0.05). At the average levels of Head Start, this implies an average Head Start rollout effect of about a three percentage-point reduction in the likelihood of adult incarceration. If one were to ascribe all of this effect to the participation margin for full-year Head Start, it would imply Head Start participation effect of 6 percentage-point reduction in the probability of ever being incarcerated. Effects of this magnitude are in line with the results from Garces, et al. (2000). Column 6 also shows that increasing K12 per-pupil spending by 10 percent (at average Head Start spending levels) reduces the likelihood of adult incarceration by 12 percentage points ($p$-value<0.05). The magnitude of this effect is in line with the estimated reductions in incarceration associated with increased schooling (Lochner and Moretti, 2003), and reductions in crime associated with attending a better school (Deming, 2011). Note, however, that this is the first paper to document a causal relationship between increased public school K12 spending and reduced risks of adult incarceration.

In sum, increases in Head Start spending improve the adult outcomes of poor children and have no effect on the outcomes of non-poor children. In contrast, increases in K12 spending improve the adult outcomes of both poor and non-poor children. These patterns also lend credibility to the research design because K12 spending increases are experienced by all children, while the direct effects of Head Start spending are experienced by only poor children. Moreover, because schools tend to be segregated by parental socioeconomic status, any indirect spillover effects of Head Start on non-Head Start enrollees will likely be experienced by poor children.

### VI.B. *Establishing the validity of the base effects*

It is important for us to establish that the variation we use in both Head Start spending and K-12 spending is exogenous. Here we summarize empirical tests that support the validity of each source of variation (for a detailed discussion of each test, see Appendix G).

*Spending Effects by Child Age*: *No confounding policies.*

One may worry that the timing of Head Start rollout or the timing of SFRs coincided with other policies that also improved adult outcomes. A test of this would be to determine whether the effects of the spending increases occur only among those who were of the appropriate age. If counties or districts adopted other policies to improve outcomes for low-income children (that were not targeted at the exact same age range as that in question), one would observe improvements for other age ranges also. To test this for Head Start, we estimated the marginal effect of the level of Head Start spending that prevailed when individuals were different ages, conditional on the level of Head Start spending when they were four years old. The marginal effects of Head Start spending on years of education and wages by age (conditional on spending at age 4), are presented in Figure 7. Higher levels of Head Start spending at age four are associated with improved adult outcomes, while the spending levels at ineligible ages (age 1 through 3 or 5 through 10) are not. This is consistent with our hypothesized mechanisms.

We conduct a similar test for K12 spending. If the spending increases we exploit operate through improved K12 education, one should see improvements for those who were between the ages of 5 and 17 when there was a SFR, but no effect for individuals from the same districts who were 18 or older at the time. To test this, we instrument for the K12 spending levels that prevailed in an individual's childhood school district when they were between the ages of 18 to 22, and we find no effect on adult outcomes. Both tests show positive effects of spending levels that prevailed when individuals were of the appropriate age, and no effect of the spending levels that prevailed when individuals were either too young or too old to be affected through the hypothesized mechanisms. This suggests that the effects are not driven by confounding policies; rather, the effects emerge through the hypothesized channels, and areas that saw spending increases were not on a pre-existing trajectory of improving outcomes. These tests support a causal interpretation of the main results.

*Instrumenting for Head Start Spending*

Because we do not instrument for Head Start spending, as we do for K12 spending, one

may still worry that the changes in Head Start spending could be related to other changes that influence outcomes. To address this, we instrument for Head Start spending with an indicator denoting whether any Head Start center had been established in one's county of birth by the year the child was age 4. In such instrumental variables models that rely only on the *timing* of the establishment of the first Head Start center in a county, the effects or Head Start spending are similar to those in Table 2. This also supports a causal interpretation of the patterns in Table 2.

*Accounting for Unobserved Family Characteristics*

Another concern one may have with the estimates is that, due to potential selective migration, the characteristics of the individuals exposed to different levels of K12 spending or Head Start spending are not the same. To rule out the possibility that our results are driven by differences across treated and untreated families, we rely on variation within families and compare the outcomes of siblings who were different ages at Head Start rollout or at the time of a court-ordered SFR, but were raised in the same household with the same parents. This approach accounts for observed and unobserved shared family characteristics that predict outcomes. We achieve this by augmenting [7] to include sibling fixed effects. In such models, the effects are similar to those in Table 2. This suggests that family selection cannot explain the main pattern of results.

*Addressing Bias due to Endogenous Mobility*

These sibling tests outlined above also address concerns regarding endogenous mobility driving the results because individuals in the same family have the same residential address. However, as an additional check on endogenous mobility, we re-estimated all models limiting the analysis sample to those who lived at their (earliest) childhood residence prior to the enactment of Head Start programs in their respective county or SFR in their district. We find similar results to Table 2, so that that endogenous residential mobility is not a source of bias in our analysis.

*Testing for Sufficient Variation to Identify the Interaction Effects*

Identification of our key parameter of interest is based on the interaction between the two policy instruments. Accordingly, our model requires that there be exogenous variation in both Head Start spending and K12 spending conditional on the other. If there were a high correlation between the two policy instruments, a model predicting both the base effects and the interaction effect could be underidentifed. We assess whether this is a problem in our setting in two ways. First, we show that correlation between Head Start spending per poor 4-year old (at age 4) and instrumented ln(K12 spending) is only 0.0844, and that conditional on our controls, there is no

association between Head Start spending and SFR-induced changes in K12 spending. Second, following Angrist and Pischke (2009), we compute a series of first-stage *F*-statistics for each set of excluded instruments, conditional on the other excluded instruments. The first-stage F-statistic on the excluded instruments for K12 spending (i.e. predicted SFR dosage times years of SFR exposure) is 51.85 and 40.86 in models without and with Head Start variables included, respectively. This shows that there is a strong first stage in predicting K12 spending, conditional on the Head Start spending levels. Additionally, in predicting the interaction between the two policies, the first-stage F-statistic on Head Start spending times SFR dosage times SFR exposure is 23.05, conditional on Head Start spending and SFR dosage times SFR exposure. This shows that there is a strong first stage in predicting the interaction between K12 spending and Head Start spending, conditional on both Head Start spending levels and our instruments for K12 spending. In sum, there is sufficient exogenous policy variation in Head Start spending and SFR-induced changes in K12 spending for the effect of each to be identified and for the interaction between the two to be identified.

### VI.C. *Evidence of Dynamic Complementarity Effects*

Before presenting the magnitudes of any complementarity effects, we first establish whether such effects exist. Specifically, in the estimation of [7], we test whether the coefficient on the interaction is positive and statistically significantly different from zero. In Table 2, for children from low-income families, the coefficient on the interaction is positive and statistically significant at the one percent level for high school graduation, years of education, and adult wage, and positive and significant at the 5 percent level for family income. For the two adverse outcomes, adult poverty and ever being incarcerated, the coefficients on the interaction terms are negative and statistically significant at the 10 percent level. That is, across all outcomes for low-income children, increases in Head Start spending raise the marginal effect of K12 spending and vice versa. In contrast, there is no such relationship for children from non-poor families (Table 3). For none of the outcomes is the coefficient on the interaction term close to statistically significant, and the signs of the coefficients across outcomes do not go in the same direction. That is, Head Start spending had no direct or indirect effect on the outcomes of non-poor children.

Because the coefficients of interaction terms can be difficult to interpret directly, we also present the interaction effects graphically. Using the regression estimates, we compute the marginal effect of increasing Head Start spending per poor four-year-old by $1,000 at each

percentile of the K12 spending distribution (between the 20th and 90th percentile).[34] We also present the 90 percent confidence interval for each estimated marginal effect. Similarly, we compute the marginal effect of increasing K12 spending by 10 percent at each percentile of the Head Start spending distribution (between the 20th and 90th percentile). The marginal effects are presented for both poor and non-poor children. Because the interactions are not significant for non-poor children, confidence intervals are provided for low-income children only.

The left panel of Figure 8 presents the estimated marginal effects of Head Start spending by the percentile of K12 spending on the likelihood of graduating from high school. If dynamic complementarity exists between early childhood and K12 spending, the plots will be upward sloping. As expected, for non-poor children (dashed line), the marginal effect of Head Start spending is flat and indistinguishable from zero at all levels of K12 spending. In contrast, for poor children (solid line), there is a clear, positive gradient. Head Start spending has small, statistically insignificant effects on children in public school districts below at the 25th percentile of per-pupil K12 spending level. However, at the 75th percentile of K12 spending, increasing Head Start spending per poor four-year-old by $1,000 increases the likelihood of graduating from high school by about 4.7 percentage points ($p$-value<0.01). The marginal effect of the same Head Start spending increase on the high school graduation rates of low-income children is about 75 percent larger in districts at the 75th percentile of the K12 spending distribution than those at the median. The right panel presents the marginal effects of increases in K12 spending at different points in the Head Start spending distribution. As expected, the marginal effect of K12 spending increases on the high school graduation rates of non-poor children is positive and is unrelated to the level of Head Start spending. However, the marginal effect of K12 spending increases on the high school graduation rates of low-income children is larger among those from childhood counties with higher levels of Head Start funding. The K12 spending effects indicate that, for poor children, a 10 percent increase in K12 spending increases the likelihood of graduating from high school by about two percentage points at the 5th percentile of the Head Start spending distribution ($p$-value>0.1) and as much as 8.5 percentage points at the 75th percentile ($p$-value<0.05). The marginal effect of the same increase in K12 spending is almost twice as large when Head Start spending is at the 75th percentile than at the 25th percentile.

[34] We do not include estimated marginal Head Start effects below the 20th percentile or above the 90th percentile of the K12 distribution because this is likely outside the range at which the marginal effects are estimated.

Figure 9 presents very similar patterns for years of completed education. Because the coefficient on the interaction term in predicting the years of education of non-poor children has a *p*-value greater than 0.5 (and is the opposite sign as that for high school graduation), we focus the discussion on low-income children. For low-income children, increasing Head Start spending per poor four-year-old by $1,000 increases the years of education completed by about 0.04 at the 25$^{th}$ percentile of K12 spending, by 0.12 years at the median of the K12 spending distribution (*p*-value<0.01), and by 0.22 at the 75$^{th}$ percentile of the K12 spending distribution (*p*-value<0.01). Looking at the marginal effect of K12 spending, increasing K12 spending by 10 percent increases years of education completed by about 0.15 at the 25$^{th}$ percentile of the Head Start spending distribution (*p*-value>0.5), by 0.24 years at the median (*p*-value<0.1), and by 0.35 at the 75$^{th}$ percentile (*p*-value<0.05). In sum, these patterns suggest important dynamic complementarity between early childhood education spending and public K12 spending for the educational outcomes of low-income children. The pattern of results indicates that, in areas with no Head Start center, increases in K12 spending may have increased educational attainment gaps. However, in areas with well-funded Head Start programs, increases in K12 spending both increased outcomes for all students and simultaneously reduced educational attainment gaps.

Commensurate with the education outcomes, there is evidence of complementarity between Head Start spending and public K12 spending in the production of adult economic outcomes for children from low-income families. Figure 10 presents the marginal effect on adult wages of each spending type by the percentile of the other spending type by childhood poverty status. For individuals from non-poor families, the marginal effect of Head Start spending is indistinguishable from zero at all levels for K12 spending levels, and a 10 percent increase in K12 spending increases the adult wage of non-poor children by about 12 percent at all levels of Head Start funding (*p*-value<0.05). This is consistent with K12 spending increases improving the outcomes of all children (including non-poor children) and Head Start spending being unrelated to the outcomes of non-poor children. We now focus on the magnitudes of the positive interaction effects for low-income children. Increasing Head Start spending per poor four-year-old by $1,000 has no appreciable effect on the adult wage at the 25$^{th}$ percentile of the K12 spending distribution, but increases the adult wage by about 2.5 percent at the median (*p*-value<0.01), and by about 5.6 percent at the 75$^{th}$ percentile (*p*-value<0.01). The dynamic complementarities are sufficiently large that the marginal effect of increases in Head Start spending on the adult wage is more than twice

as large when K12 spending is at the 75th percentile than at the median.

We now turn to the marginal effect of increases in K12 spending. Increasing K12 spending by 10 percent increases the adult wages of poor children by about 7 percent at the 25th percentile of Head Start spending ($p$-value>0.1), by 8.5 percent at the median ($p$-value<0.05), and by 13 percent at the 75th percentile ($p$-value<0.05). Similar to the education outcomes, the marginal effect of increases in K12 spending (on the adult wages of poor children) is about twice as large when Head Start spending is at the 75th percentile than at the 25th percentile. The pattern of results for adult family income is similar to those for adult wages and is presented in Appendix H.

We also examine effects on the annual incidence of poverty in adulthood. A family is poor if their income-to-needs ratio is below the federally-determined threshold for poverty. Furthermore, while adult poverty is *related* to family income, it is a measure of hardship. The marginal effects on the annual incidence of adult poverty are presented in Figure 11. For children from non-poor families, neither spending increases in K12 spending nor increases in Head Start spending affect the likelihood of adulthood poverty. We therefore focus our discussion on effects for poor children. For poor children, increasing Head Start spending per poor four-year-old by $1,000 reduces the annual incidence of poverty in adulthood by about 0.6 percentage points at the 25th percentile of the K12 spending distribution ($p$-value>0.1). Consistent with dynamic complementarity, that same increase in Head Start spending reduced adult poverty by 1.2 percentage points for those from K12 school districts at the median of the K12 spending distribution ($p$-value<0.05), and 1.7 percentage points at the 75th percentile ($p$-value<0.05). Relative to the baseline poverty rate, this marginal effect at the 75th percentile of the K12 spending distribution represents about a 10 percent reduction. The differences in the marginal effect of K12 spending on adult poverty by Head Start spending are more noisily estimated (even though the interaction term is statistically significant at the 5 percent level). However, the general pattern of larger benefits to K12 spending at higher levels of Head Start spending is maintained.

The final outcome we examine is the likelihood of ever being incarcerated. The marginal effects are presented in Figure 12. As with adulthood poverty, we focus our discussion on effects for children from low-income families. As with other outcomes, there is evidence of strong dynamic complementarity between Head Start spending and public K12 spending. For poor children, increasing Head Start spending by $1,000 had no effect on incarceration among children in public school districts at the 25th percentile of K12 spending distribution, but reduced it by 1.6

percentage points for students in districts at the median ($p$-value<0.05), and reduced it by 2.2 percentage points for students in districts at the 75$^{th}$ percentile ($p$-value<0.05). Looking to the effect of K12 spending on the likelihood of being incarcerated, the marginal effects are larger at higher levels of Head Start spending (notice the difference in scale), but the K12 spending effects are rather imprecise. However, the point estimates indicate that a 10 percent increase in K12 per-pupil spending reduces the likelihood of being incarcerated by 10 percentage points with no Head Start spending ($p$-value<0.05), by 13 percentage points at the median of the Head Start spending distribution ($p$-value<0.05), and 15 percentage points at the 75$^{th}$ percentile ($p$-value<0.05). As with the other adult outcomes, the reduction in the lifetime risks of incarceration associated with improvements in access to early education is larger when there are greater subsequent K12 school investments and *vice versa*.

### VI.D. *Is Parenting Quality Part of the Story?*

In Section II, we posited that the Head Start effects are driven by the components targeted to children. However, because parent counseling was a component of Head Start, it is possible that these dynamic complementarities emerge through improvements in parenting quality. Because we have data on siblings who shared the same parents, we can test for improvements in parenting quality directly. Specifically, we can use only the sample of older siblings who were not themselves exposed to Head Start, and test whether those with younger siblings who were exposed to Head Start have improved outcomes. If improvements in parenting quality is a part of the story, the older siblings of exposed younger siblings should have better outcomes than the older siblings of unexposed younger siblings. However, if the Head Start effects are driven by the services provided to the children, there should be no effect. In these models (Appendix I), we find older siblings are unaffected by Head Start exposure of the younger sibling. This suggests that parenting quality is not part of the story and that our Head Start spending effects reflect real investments in the human capital of low-income children.

### VI.E. *Are the Complementarity Effects Driven by Other Coincident Policies?*

Even though our estimation equations control for several coincident polices directly, one may worry that our main results are driven by some complementarity between K12 spending and some other policy. To test for this directly, we augment our main model in equation [7] to also include (a) interactions between food stamp spending in one's county between ages 0 to 4 with K12 spending, and (b) county level spending on Medicaid between ages 0 and 4 interacted with

K12 spending. In these models, the point estimates on the interaction between Head Start spending and K12 spending are virtually unchanged.[35] This provides further evidence that our estimated effects are not confounded by dynamic complementarities with other policies.

## VII. BENEFIT-COST CONSIDERATIONS: PUTTING THE MAGNITUDES IN PERSPECTIVE

It is helpful to consider how the presence of dynamic complementarity affects the optimal allocation of resources to the K12 system versus to early childhood education (for poor children). If average outcomes are maximized, the marginal dollar spent on Head Start will yield the same effect on outcomes as an equivalent expenditure on K12 education. It is helpful to define some parameters. The proportion of poor children in a county is *p*. The marginal effect of a $1,000 increase in Head Start spending on low-income children is $\delta_{HS,\,poor}$, and the marginal effect of a $1,000 increase in Head Start spending on non-poor children is $\delta_{HS,non}$. As such, the average effect of a 1,000 increase in Head Start spending in a county is [8] below.

[8] $$\pi_{HS} = p\delta_{HS,poor} + (1-p)\delta_{HS,non}$$

To equate the marginal effects of spending on Head Start to that of spending on the K12 system, we need to define the change in K12 spending that would lead to the same expenditure as an increase of $1,000 in Head Start spending per poor-four-year old. During our sample period, K12 spending was roughly $4,000 per student per year *on average*. Assuming a 7 percent interest rate, spending $4,000 for 12 years is equivalent to $34,000 in present value terms. Thus, an equivalent expenditure at the *student* level would be a 1,000/34,000=2.94 percent increase in K12 spending. However, this increase is only spent on the poor children so that the equivalent spending increase in percentage terms for all children in a *county* is 2.94**p*. $\delta_{K12,poor}$ and $\delta_{K12,non}$ are the effect of increasing K12 spending by one percent on poor and non-poor children, respectively. The marginal effect of the equivalent increase in K12 spending on the average child in the county is therefore

[9] $$\pi_{K12} = (p\delta_{K12,poor} + (1-p)\delta_{K12,non})(2.94\text{p})$$

The ratio between these two equations $\pi_{HS}/\pi_{K12}$ is the relative effectiveness of spending $1,000 per poor four year old on Head Start and spending the same amount across all children from that same cohort in the county in the K12 system. Because $\delta_{HS,non} = 0$, this ratio simplifies to [10].

[10] $$\frac{\pi_{HS}}{\pi_{K12}} = \frac{p\delta_{HS,poor}}{(p\delta_{K12,poor}+(1-p)\delta_{K12,non})(2.94\text{p})} = \frac{\delta_{HS,poor}/(2.94)}{p(\delta_{K12,poor}-\delta_{K12,non})+\delta_{K12,non}}$$

[35] Results from this additional model are available upon request.

The relative marginal effect of Head Start spending and K12 is a function of the poverty rate *p* as long as $\delta_{K12,poor} \neq \delta_{K12,non}$. Specifically, if $\delta_{K12,poor} > \delta_{K12,non}$ , then this ratio is falling in *p*, and if $\delta_{K12,poor} < \delta_{K12,non}$ this ratio is increasing in *p*. Intuitively, if non-poor children are more responsive than poor children to increases in K12 spending (i.e. $\delta_{K12,poor} < \delta_{K12,non}$), then the marginal benefit of increased K12 spending declines with the poverty rate so that the *relative* effectiveness of Head Start spending increases with the poverty rate. The converse is also true.

To show the relationship between this ratio and the poverty rate at the average level of K12 spending, in Figure 13 we plot this ratio against the poverty rate, where this ratio is evaluated at the mean levels of K12 and Head Start spending in the data (solid gray line). This ratio of marginal effects is plotted against the poverty rate for years of education on the left and adult wages on the right. The model is linear in Head Start spending but linear in the *log* of K12 spending. Therefore, the marginal effect of K12 spending will fall relative to that for Head Start at higher levels of K12 spending even without any dynamic complementarity. To show this relationship, we plot $\pi_{HS}/\pi_{K12}$ against the poverty rate where the present value is evaluated at above average K12 spending of $5,000 per-pupil and the interaction term is assumed to be zero. This is depicted by the dashed grey line. The difference between the solid grey line and the dashed grey line depicts the change in this ratio at above average K12 spending assuming that there is no dynamic complementarity. To illustrate how dynamic complementarity affects this ratio, we plot $\pi_{HS}/\pi_{K12}$ against the poverty rate where the present value is evaluated at above average K12 spending of $5,000 per-pupil, but where the interaction term is non-zero (i.e., using the empirical estimate of the interaction term from Table 2). Accordingly, the difference between the dashed grey line and the solid black line depicts the change in this ratio at above average K12 spending due only to dynamic complementarity.

The ratio for years of education is on the left and for adult wage on the right. For years of education, at all poverty levels, the ratio of the marginal effect of spending on Head Start to that of the equivalent in K12 spending is above one at average levels of Head Start and K12 spending. This suggests that at average spending levels between 1965 and 1980 one dollar spent on Head Start lead to a larger marginal improvement in the average years of education in the county than spending one dollar on K12 education. As expected, when K12 spending was above average, the ratio is even higher at all poverty levels even without any dynamic complementarity. More precisely, the ratio increases from about 1.5 to 1.9 when K12 spending is 1,000 above the average

(ignoring the contribution of complementarity). Taking dynamic complementarity into account, the ratio at above-average K12 spending levels increases considerably to about 4. That is, if one accounts for dynamic complementarity, for a district that spent $5,000 per K12 pupil and $4,000 per poor four year old during our sample period, each marginal dollar spending on Head Start had four times the effect of that same dollar spent on K12 education. Importantly, this holds for counties at all poverty levels. The figure on the right for wages presents similar patterns. At average spending levels, the ratio is below 1 (about 0.6), and at above-average spending levels ($5,000 per-pupil) not accounting for dynamic complementarity, the ratio is about 0.75. After accounting for dynamic complementarity, the ratio increases to about 2.1 at above average spending levels. For adult wages, the ratio is below 1 when does not account for dynamic complementarity, but is well above 1 when one accounts for dynamic complementarity.

In essence, these patterns support the idea that, when such dynamic complementarities exist between early and late human capital investments, there may be no equity-efficiency tradeoff when shifting resources toward compensatory early education programs (Cunha and Heckman, 2007). More specifically, our estimates indicate that, for a district that spent $5,000 per-pupil (about 20 percent above the average K12 spending level), the marginal dollar spent on Head Start led to between 2 and 4 times the improvement in adult outcomes as that spent on K12 education. Accordingly, at such spending levels, one could redistribute money from the K12 system towards Head Start and have *both* better average outcomes and a more equitable distribution of adult outcomes. Overall, the patterns in Figure 13 suggests that during our sample period, the marginal dollar had a roughly equal effect on adult outcomes overall at levels close to the national averages that prevailed at that time. The patterns also indicate that when resources are allocated efficiently, localities with higher levels of Head Start spending should have higher levels of K12 spending and vice versa. Empirically, the correlation between per-pupil spending and Head Start spending is roughly 0.35. This implies that, in general, localities may be taking advantage of these complementarities, but that further optimization is likely possible.

### VII.A Summary & Conclusions

This study provides new evidence on the life-cycle effects of Head Start spending and K12 school spending for low-income children. That is, we explore dynamic complementarities between human capital investments made in pre-school and those that subsequently occur in the K12

system. In particular, we use children's differential exposure to Head Start spending (at age 4) and court-ordered school finance reforms (between the ages 5 through 17), depending on place and year of birth, to examine whether the marginal effect of Head Start spending on children's adult outcomes are larger among individuals who were subsequently exposed to SFR-induced K12 spending increases. We present extensive tests to document that the policy-induced variation in Head Start spending and K12 public school spending we exploit is unrelated to other childhood family, community, or policy changes.

For non-poor children, SFR-induced K12 spending increases led to significant improvements in educational and economic outcomes, while increases in Head Start spending had no effect. However, for low-income children, both Head Start spending increases and SFR-induced K12 spending increases led to significant improvements in educational outcomes, economic outcomes, and reductions in the likelihood of incarceration. Importantly, the long-run effects of increases in Head Start spending are amplified when followed by attending schools that experienced SFR-induced increases in K12 per-pupil spending. Across all the outcomes, the marginal effect of the same increase in Head Start spending was more than twice as large for students born in K12 school districts that spent at the 75$^{th}$ percentile of the distribution than those born in K12 school districts that spent at the 25$^{th}$ percentile. Similarly, the benefits of K12 school-spending increases on adult outcomes were larger among poor children who were exposed higher levels to Head Start spending during their pre-school years. For poor children, the combined benefits of growing up in districts/counties with *both* greater Head Start spending and K12 per-pupil spending are significantly greater than the sum of the parts (i.e., the independent effects of increases in Head Start or K12 school spending in isolation).

There are two important caveats to our work. First, because the counterfactual childcare and pediatric care may be better today than in the late 1960s and 70s, the marginal effect of Head Start may be smaller today than in the earlier period that we study.[36] Second, public school spending levels during the period we study were lower than current levels. If school spending exhibits diminishing marginal product, the effects presented here may be larger than one would

[36] In the early period of Head Start, most poor children would have received home care, while today, as many as one-third of Head Start participants may have attended some other form of formal childcare (Kline and Walters, 2016; Feller et al., 2016). The proportion of three- and four-year-olds in school has increased from roughly 10 percent in 1964 to almost 40 percent by 1995 (source: US Census Bureau, CPS October Supplement, 1964-2010; see Figure 1). Also, while most poor children currently receive pediatric care through Medicaid and SCHIP, during the period under study many children would only have received such care through Head Start.

observe with similar spending increases today. These caveats do not minimize the importance of the findings or their profound implications for policy. However, they do suggest that the contemporary magnitude of the effects may be smaller than those we present here.

The cumulative nature of skill development is likely responsible for the pattern of results. Our findings highlight the importance of modeling early and later educational investments jointly and may provide an explanation for some disparate results on the effects of Head Start. Indeed, our finding that the long-run effects of Head Start are larger among individuals who attended better resourced schools may provide an explanation for why Head Start may have been more successful for more socioeconomically-advantaged populations (Currie and Duncan, 1995) and why there is a fade out of the effects of Head Start on test scores as students age (Currie and Duncan, 2000). The key policy implication of our findings is that human capital investments made in and sustained throughout child developmental stages (pre-school; elementary/middle school; adolescence) may yield greater returns than separate, isolated, short-lived reforms not sustained beyond the year in which they are implemented. The findings point to the critical role early-life investments can play in narrowing long-run gaps in well-being, but they also highlight the importance of sustained investments in the skills of disadvantaged youth.

## References


Aizer Anna, and and Cunha, Flavio (2012) The Production of Human Capital: Endowments, Investments and Fertility" NBER Working Paper No. 18429.

Almond, Douglas and Janet Currie (2010). Handbook of Labor Economics, chapter "Human Capital Development Before Age Five", edited by Orley Ashenfelter and David Card.

Almond, Douglas, Hilary Hoynes, and Diane Whitmore Schanzenbach (2011). "Inside the War on Poverty: The Impact of the Food Stamp Program on Birth Outcomes". Review of Economics and Statistics, Vol. 93, No. 2: 387-403.

Angrist, Joshua D. and Alan B. Krueger. 1992. "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples." Journal of the American Statistical Association 87(418):328–336.

Ben-Shalom, Yonatan, Robert Moffitt, and John Karl Scholz. An Assessment of the Effectiveness of Anti Poverty Programs in the United States. No. 17042. National Bureau of Economic Research, Inc, 2011.Becketti, Sean, William Gould, Lee Lillard, and Finis Welch. 1988. The PSID after Fourteen Years: an Evaluation. *Journal of Labor Economics* 6, no. 4: 472-92.

Blau, David, and Janet Currie. "Pre-school, day care, and after-school care: who's minding the kids?." *Handbook of the Economics of Education* 2 (2006): 1163-1278.

Bjorklund, A. and M. Jantti, Intergenerational Income Mobility in Sweden Compared to the United States, American Economic Review, 87, 5, December 1997.

Card, David, and Krueger, Alan B. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." Journal of Political Economy, February 1992, 100(1), pp. 1-40.

Card, D., and A. A. Payne. "School Finance Reform, the Distribution of School Spending, and the Distribution of Student Test Scores." Journal of Public Economics, 83(1) (2002): 49–82.

Carneiro, Pedro, and Rita Ginja. 2014. "Long-Term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start." *American Economic Journal: Economic Policy*, 6(4): 135-73.

Cascio, Elizabeth (2009). "Do Investments in Universal Early Education Pay Off? Long-term Effects of Introducing Kindergartens into Public Schools". NBER working paper #14951.

Chay, K. Y., J. Guryan, and B. Mazumder. "Birth Cohort and the Black-White Achievement Gap: The Roles of Access and Health Soon After Birth." NBER Working Paper No. 15078 (2009).

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schazenbach, and Danny Yang (2011). "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence From Project STAR". Quarterly Journal of Economics (2011) 126 (4): 1593-1660.

Cohodes, Sarah., Grossman, Daniel,. Kleiner, Samuel., Lovenheim, Michael 2015 "The Effect of Child Health Insurance Access on Schooling: Evidence from Public Insurance Expansions" Journal of Human Resources August 1, 2016 51:556-588

Currie, Janet and Duncan Thomas (1995). "Does Head Start make a difference?" American Economic Review 85(3): 341-364.

Currie, Janet and Mathew Neidell (2007). "Getting Inside the "Black Box" of Head Start Quality: What Matters and What Doesn't". Economics of Education Review 26(1): 83-99.

Currie, Janet and Duncan Thomas (2000). "School Quality and the Long-term Effects of Head Start." Journal of Human Resources 35(4): 754-774.

Currie, Janet and Yelowitz, Aaron, (2000), Are public housing projects good for kids?, *Journal of Public Economics*, **75**, issue 1, p. 99-124

Cunha, Flavio & James Heckman, 2007. "The Technology of Skill Formation," American Economic Review, American Economic Association, vol. 97(2), pages 31-47, May.

Cunha, Flavio, James J.Heckman, Lance J.Lochner, and Dimitriy V. Masterov. 2006. "Interpreting the Evidence on Life Cycle Skill Formation." In Handbook of the Economics of Education, ed. Eric A. Hanushek and Frank Welch, 697–812. Amsterdam: North-Holland: Elsevier.

Cutler, David and Adriana Lleras-Muney (2008). "Education and Health: Evaluating theories and evidence". Published in Making Americans Healthier: Social and Economics Policy as Health Policy, Robert F. Schoeni, James S. House, George Kaplan and Harold Pollack, editors, New York: Russell Sage Foundation.

Dee, Thomas S. and William N. Evans *Journal of Labor Economics* Vol. 21, No. 1 (January 2003), pp. 178-209.

Deming, David. "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start." American Economic Journal: Applied Economics 1, no. 3 (July 1, 2009): 111–34.

Deming DJ. Better Schools, Less Crime?. Quarterly Journal of Economics. 2011;126 (4) :2063-2115.

Dobbie, Will, and Roland G. Fryer, 2011, "Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone," American Economic Journal: Applied Economics, Vol. 3, No. 3, July, pp. 158–187.

Lipsey, M. W., Farran, D.C., & Hofer, K. G., (2015). A Randomized Control Trial of the Effects of a Statewide Voluntary Prekindergarten Program on Children's Skills and Behaviors through Third Grade (Research Report). Nashville, TN: Vanderbilt University, Peabody Research Institute.

Feller, Avi, Todd Grindal, Luke Miratrix, and Lindsay Page, 2016. "Compared to what? Variation in the impacts of early childhood education by alternative care type" *Annals of Applied Statistics*. 2016. 110(3): 1245-1285.

Fitzgerald, John, Peter Gottschalk, and Robert Moffitt. 1998. An Analysis of the Impact of Sample Attrition on the Second Generation of Respondents in the Michigan Panel Study of Income Dynamics. *The Journal of Human Resources* 33, no. 2: 300-344.

Fitzgerald, John, Peter Gottschalk, and Robert Moffitt. 1998. An Analysis of Sample Attrition in Panel Data. *The Journal of Human Resources* 33, no. 2: 251-99.

Fryer , Roland G. & Steven D. Levitt, 2004. "Understanding the Black-White Test Score Gap in the First Two Years of School," The Review of Economics and Statistics, MIT Press, vol. 86(2), pages 447-464, 06.

Fryer, Roland, and Steven Levitt (2006). "The Black-White Test Score Gap Through Third Grade." American Law and Economic Review, 8(2): 249-281.

Fuerst, J.S. and D. Fuerst. (1993). Chicago Experience with an Early Childhood Program: The Special Case of the Child Parent Center Program. URBAN EDUCATION 28(1, Apr): 69-96. EJ 463 446.

Garces, Eliana, Duncan Thomas and Janet Currie (2002). "Longer-term Effects of Head Start." American Economic Review 92(4): 999-1012.

García, Jorge Luis and Gallegos, Sebastian, Dynamic Complementarity or Substitutability? Parental Investment and Childcare in the Production of Early Human Capital (February 1, 2017). Available at SSRN: https://ssrn.com/abstract=2910167 or http://dx.doi.org/10.2139/ssrn.2910167

Gibbs, Chloe and Ludwig, Jens and Miller, Douglas L., Does Head Start Do Any Lasting Good? (September 2011). NBER Working Paper No. w17452.

Gilraine (2016) "School Accountability and the Dynamics of Human Capital Formation" University of Toronto Mimeo.

Hanushek, Eric A. Spending on Schools. Stanford, Calif.: Hoover Press, 2001. http://hanushek.stanford.edu/sites/default/files/publications/Hanushek%202001%20PrimerAmerEduc.pdf.

_____., "School Resources and Student Performance," in Gary Burtless, ed., Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success,
Brookings Institution, Washington, D.C., 1996, pp. 43-73

_____. "The Economics of Schooling: Production and Efficiency in Public Schools." Journal of Economic Literature, 1986, 1141–77.

Heckman, James J. (2008). "Schools, Skills, and Synapses," Economic Inquiry, vol. 46(3), pages 289-324.

_____. (2007). The economics, technology, and neuroscience of human capability formation. PNAS, 104(33):13250–13255.

Heckman , James J. and Stefano Mosso "The Economics of Human Development and Social Mobility" Annual Review of Economics 2014. 6:689–733.

Heckman, J., R. Pinto, and P. Savelyev. "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." American Economic Review (forthcoming).

Hoxby, C. M. "All School Finance Equalizations Are Not Created Equal." The Quarterly Journal of Economics (2001): 1189–1231.

Hoynes, Hilary, Marianne Page, and Ann Stevens (2011). "Can Targeted Transfers Improve Birth Outcomes? Evidence from the Introduction of the WIC Program", Journal of Public Economics, 95: 813–827.

Hilary Hoynes & Diane Whitmore Schanzenbach & Douglas Almond, 2016. "Long-Run Impacts of Childhood Access to the Safety Net," American Economic Review, vol 106(4), pages 903-934.

Jackson, Kirabo, Rucker C. Johnson, Claudia Persico (2016). “The Effects of School Spending on Educational & Economic Outcomes: Evidence from School Finance Reforms”. The Quarterly Journal of Economics, vol 131(1), pages 157-218.

_____. The Effect of School Finance Reforms on the Distribution of Spending, Academic Achievement, and Adult Outcomes. NBER Working Paper 20118, May (2014).

Jackson, C. Kirabo. 2012, "Non-Cognitive Ability, Test Scores, and Teacher Quality: Evidence from 9th Grade Teachers in North Carolina", NBER Working Paper No. 18624

Jackson, C. Kirabo. (forthcoming) "What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes" *Journal of Political Economy.*

Johnson, Rucker C. (2015). “Follow the Money: School Spending from Title I to Adult Earnings”. Special edited volume, ESEA at 50, forthcoming in The Russell Sage Foundation Journal of the Social Sciences.

_____. 2015. "Can Schools Level the Intergenerational Playing Field? Lessons from Equal Educational Opportunity Policies". Forthcoming in Federal Reserve volume on mobility.

_____. “Long-run Impacts of School Desegregation & School Quality on Adult Attainments.”NBER Working Paper No. 16664 (2011) , updated August 2015.

_____. (2010). “The Health Returns of Education Policies: From Preschool to High School & Beyond.” American Economic Review Papers and Proceedings (May), 100(2): 188-94.

Kline, Patrick & Christopher R. Walters, 2016. "Evaluating Public Programs with Close Substitutes: The Case of Head Start," The Quarterly Journal of Economics, vol 131(4), pages 1795-1848.

Knudsen, Eric I., James J. Heckman, Judy Cameron and Jack P. Shonkoff. 2006. “Economic, Neurobiological and Behavioral Perspectives on Building America’s Future Workforce.” Proceedings of the National Academy of Sciences 103 (27): 10155–62.

Levine, Phillip B. & Diane Schanzenbach, 2009. "The Impact of Children's Public Health Insurance Expansions on Educational Outcomes," Forum for Health Economics & Policy, Berkeley Electronic Press, vol. 12(1).

Lochner, Lance, and Enrico Moretti. 2004. "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports." *American Economic Review*, 94(1): 155-189.

Loeb, Susanna, and John Bound. “The Effect of Measured School Inputs on Academic Achievement: Evidence from the 1920s, 1930s and 1940s Birth Cohorts.” The Review of Economics and Statistics 78, no. 4 (November 1, 1996): 653–64. doi:10.2307/2109952.

Lubotsky, Darren and Robert Kaestner, “Do ‘Skills Beget Skills’? Evidence on Dynamic Complementarities in Cognitive and Non-Cognitive Skills in Childhood,” Economics of Education Review Volume 53, August 2016, Pages 194–206.

Ludwig, Jens, and Douglas L. Miller. “Does Head Start Improve Children’s Life Chances? Evidence from a Regression Discontinuity Design.” The Quarterly Journal of Economics 122, no. 1 (February 1, 2007): 159–208. doi:10.1162/qjec.122.1.159.

Magnuson, Katherine A. and Waldfogel, Jane “Early Childhood Care and Education: Effects on Ethnic and Racial Gaps in School Readiness” Future of Children VOL. 15 NO. 1 SPRING 2005.

Malamud, Ofer., Pop-Eleches, Christian., and Urquiola, Miguel “Interactions Between Family and School Environments: Evidence on Dynamic Complementarities?” (2016) NBER Working Paper No. 22112.

Murray, S. E., W. N. Evans, and R. M. Schwab. "Education-Finance Reform and the Distribution of Education Resources." American Economic Review, 88(4) (1998): 798–812.

Newport, Elissa L. (1990). “Maturational constraints on language learning.” *Cognitive Science* 14: 11-28.

Oden, S., Schweinhart, L. J., Weikart, D. P., Marcus, S. M., & Xie, Y. (2000). Into adulthood: A study of the effects of Head Start. Ypsilanti, MI: High/Scope Press.

Pinker, Steven. (1994). *The Language Instinct*. New York, NY: Harper Perennial Modern Classics.

Rossin-Slater Maya, and Wust Miriam “What is the Added Value of Preschool? Long-Term Impacts and Interactions with a Health Intervention” (2016) †University of California at Santa Barbara mimeo.

Sacerdote, Bruce. 2014. “Experimental and Quasi-experimental Analysis of Peer Effects: Two Steps Forward?” *Annual Review of Economics*, volume 6: 253-272.

U.S. Department of Education, National Center for Education Statistics. Public School Finance Programs of the United States and Canada: 1998–99. NCES 2001–309; Compilers Catherine C. Sielke, John Dayton, and C. Thomas Holmes, of the University of Georgia and Anne L. Jefferson of the University of Ottawa. William J. Fowler, Jr., Project Officer. Washington, DC: 2001.

Zigler, Edward & J. Valentine (Eds.). 1979. *Project Head Start: A legacy of the war on poverty*. New York: Free Press.

Zigler, Edward**.** 2010. “Putting the National Head Start Impact Study into a Proper Perspective.” *National Head Start Association* (*NHSA*) 13 (1): 1–6.

Zigler, Edward, Gilliam, & W., Barnett, W.S. (Eds.) (2011). Current debates and issues in prekindergarten education. Baltimore, MD: Paul H. Brookes.

**Table 1:**
*Summary Statistics of the Analytic Dataset*

| | All (N=13,381) | Poor Child (N=5,623) | Non-Poor Child (N=7,758) |
|---|---|---|---|
| Adult Outcomes: | | | |
| High School Graduate | 0.85 | 0.71 | 0.89 |
| Years of Education | 13.29 | 12.29 | 13.61 |
| Ln(Wages), at age 30 | 2.49 | 2.24 | 2.56 |
| Adult Family Income, at age 30 | $48,655 | $35,372 | $52,448 |
| In Poverty, at age 30 | 0.08 | 0.18 | 0.05 |
| Ever Incarcerated | 0.05 | 0.08 | 0.04 |
| Age (range: 20-50) | 30.8 | 30.3 | 31.0 |
| Year born (range: 1950-1976) | 1962 | 1962 | 1962 |
| Female | 0.44 | 0.43 | 0.44 |
| White | 0.87 | 0.66 | 0.93 |
| Childhood school variables: | | | |
| Any Head Start Center in county, age 4 | 0.33 | 0.33 | 0.34 |
| Post rollout: Head Start spending per poor 4-year old, age 4 | $4,103 | $4,204 | $4,072 |
| Child attended Head Start* | 0.04 | 0.19 | 0.02 |
| Child attended any pre-school program | 0.23 | 0.31 | 0.23 |
| School District Per-pupil spending (avg, ages 5-17) | $4,277 | $4,013 | $4,363 |
| Any court-ordered school finance reform, age 5-17 | 0.13 | 0.11 | 0.14 |
| Cond'l on any: # of exposure yrs to school finance reform | 7.37 | 6.90 | 7.50 |
| 1960 District Poverty Rate (%) | 21.52 | 28.25 | 19.35 |
| Childhood family variables: | | | |
| Income (avg, ages 12-17) | $54,488 | $22,520 | $65,130 |
| Income-to-needs ratio (avg, ages 12-17) | 3.05 | 1.31 | 3.62 |
| Mother's years of education | 11.84 | 10.61 | 12.24 |
| Father's years of education | 11.82 | 10.04 | 12.36 |
| Born into two-parent family | 0.90 | 0.74 | 0.95 |
| Low birth weight (<5.5 pounds) | 0.07 | 0.07 | 0.07 |

Note: All descriptive statistics are sample weighted to produce nationally-representative estimates of means. Dollars are CPI-U deflated in real 2000 $. Poor children are those whose parents were in the bottom quartile of the income distribution (approximately 80% of whom were below the poverty line). Analysis sample includes 13,381 individuals (191,613 person-year observations ages 20-50), from 4,684 childhood families, 1,431 school districts, 1,070 childhood counties and all 50 states.

*Child-specific pre-K attendance & Head Start program participation info collected retrospectively in 1995 survey IW.

**Table 2:**

*2SLS-Difference-in-Difference Estimates of Early and K12 education Spending on Adult Outcomes*

| | Children from Poor Households | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| | Prob(High School Grad) | Years of Education | Ln(Wage), ages 20-50 | Ln(Family Income), ages 20-50 | Poverty, ages 20-50 | Prob(Ever Incarcerated) |
| Head Start Spending$_{(age\ 4)}$/1000 | 0.0198*** | 0.0967*** | 0.0193*** | 0.0359*** | -0.009*** | -0.007541** |
| | (0.0039) | (0.0191) | (0.0050) | (0.0075) | (0.0028) | (0.003285) |
| Ln(K12 Per-pupil Spending)$_{(age\ 5\text{-}17)}$ | 0.595** | 2.3733* | 0.9271* | 0.4476 | -0.01846 | -1.2751** |
| | (0.2370) | (1.3883) | (0.5277) | (1.0880) | (0.3331) | (0.5804) |
| Interaction | 0.118*** | 0.4936*** | 0.150*** | 0.1120** | -0.03302* | -0.06045* |
| | (0.0455) | (0.1424) | (0.0330) | (0.0534) | (0.0187) | (0.03597) |
| Number of Person-year Observations | | | 49,282 | 54,064 | 54,309 | |
| Number of Children | 5,385 | 5,378 | 5,378 | 5,378 | 5,378 | 4,317 |

Robust standard errors in parentheses (clustered at childhood state level)

*** p<0.01, ** p<0.05, * p<0.1

Data: PSID geocode Data (1968-2013), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 whose parents were in the bottom quartile of the income distribution, and who have been followed into adulthood.

Models: Head Start spending per poor 4-year old in the county is centered around $4,230 (and measured in 000s) and instrumented ln(school district per-pupil spending during ages 5-17) is centered around 1.6, to facilitate interpretation of the main effects evaluated at roughly the respective means; the average SFR-induced increase in school-age spending is about 10%. Results are based on 2SLS-Difference-in-Difference models that include: parent's relative rank in income distribution, school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, rollout of "War on Poverty" & related safety-net programs (community health centers, food stamps, medicaid, AFDC, UI, Title-I (average during childhood yrs)), timing of state-funded Kindergarten introduction and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends; and controls for childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). The first-stage model include as predictors the school-age years of exposure to school finance reform interacted with the quartile of the respective school district's predicted reform-induced change in school spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories. There exists a significant first-stage for all endogenous regressors. See Appendix Table G7 for the partial fist-stage F-statistics.

**Table 3:**

*2SLS-Difference-in-Difference Estimates of Early and K12 education Spending on Adult Outcomes*

| | Children from Non-Poor Households | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| | Prob(High School Grad) | Years of Education | Ln(Wage), ages 20-50 | Ln(Family Income), ages 20-50 | Poverty, ages 20-50 | Prob(Ever Incarcerated) |
| Head Start Spending$_{(age\ 4)}$/1000 | 0.003162 | 0.02351 | 0.006178 | 0.001183 | 1.64E-04 | -0.001766 |
| | (0.0029) | (0.0159) | (0.0052) | (0.0096) | (0.0014) | (0.001613) |
| Ln(K12 Per-pupil Spending)$_{(age\ 5\text{-}17)}$ | 0.3159 | 1.9886 | 1.1730** | 1.1293* | -0.0192 | 0.1775 |
| | (0.2891) | (1.6299) | (0.4565) | (0.5952) | (0.0994) | (0.1372) |
| Interaction | 0.02439 | -0.1684 | 0.03147 | -0.03185 | 0.01019 | -0.01048 |
| | (0.0405) | (0.2551) | (0.0416) | (0.0891) | (0.0131) | (0.01284) |
| Number of Person-year Observations | | | 79,239 | 81,598 | 81,737 | |
| Number of Children | 7,718 | 7,729 | 7,729 | 7,729 | 7,729 | 4,364 |

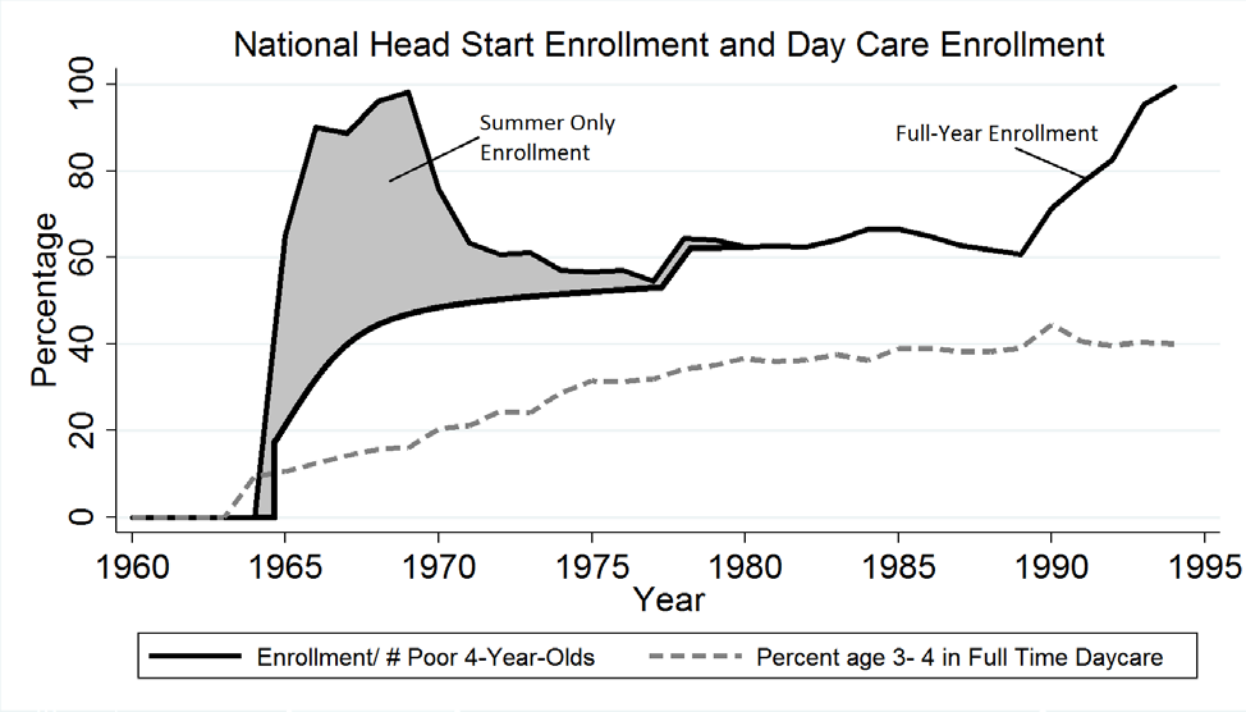
Robust standard errors in parentheses (clustered at childhood state level)

*** p<0.01, ** p<0.05, * p<0.1

Data: PSID geocode Data (1968-2013), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 whose parents were in the bottom quartile of the income distribution, and who have been followed into adulthood.

Models: Head Start spending per poor 4-year old in the county is centered around $4,230 (and measured in 000s) and instrumented ln(school district per-pupil spending during ages 5-17) is centered around 1.6, to facilitate interpretation of the main effects evaluated at roughly the respective means; the average SFR-induced increase in school-age spending is about 10%. Results are based on 2SLS-Difference-in-Difference models that include: parent's relative rank in income distribution, school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, rollout of "War on Poverty" & related safety-net programs (community health centers, food stamps, medicaid, AFDC, UI, Title-I (average during childhood yrs)), timing of state-funded Kindergarten introduction and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends; and controls for childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). The first-stage model include as predictors the school-age years of exposure to school finance reform interacted with the quartile of the respective school district's predicted reform-induced change in school spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories. There exists a significant first-stage for all endogenous regressors. See Appendix Table G7 for the partial fist-stage F-statistics.

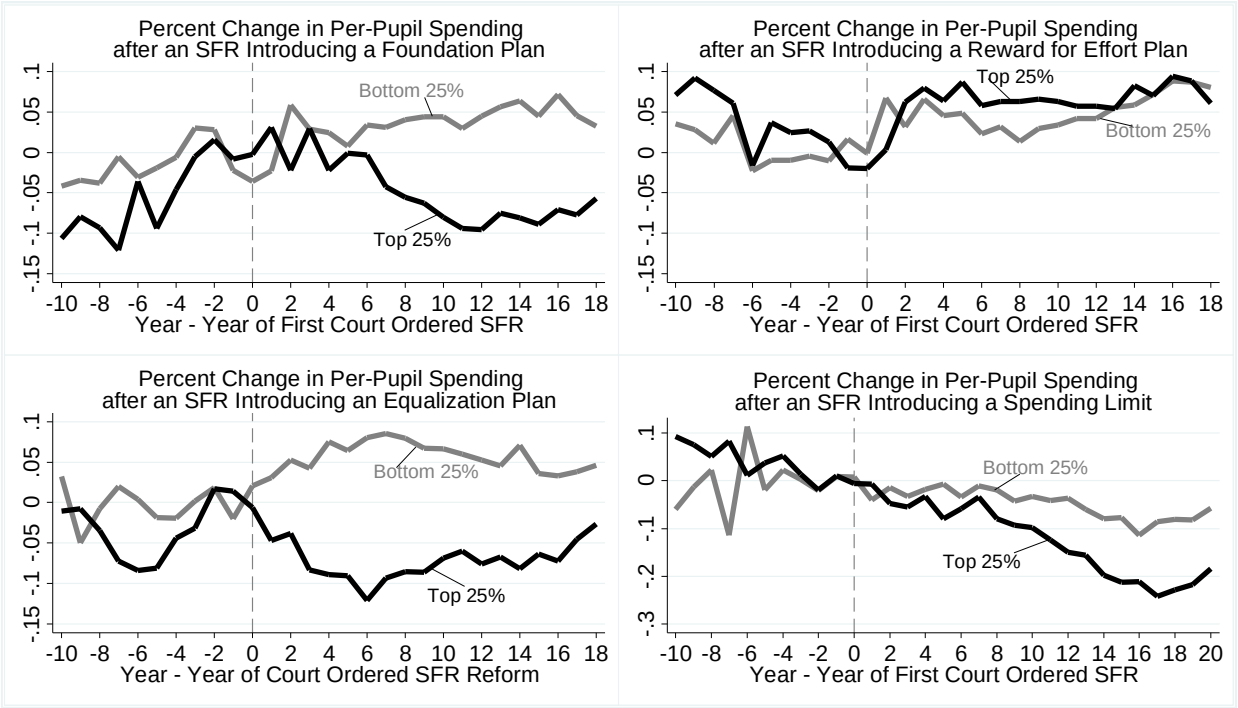
**Figure 1:**
*National Enrollment in Head Start Over Time*




Notes: The solid black line in this figure plots the national enrollment in Head Start divided by the number of poor four-year-olds in the nation in that same calendar year. Head Start enrollment figures are from the Head Start fact sheet (link: https://eclkc.ohs.acf.hhs.gov/hslc/data/factsheets/2015-hs-program-factsheet.html) and the national counts of poor four-year-olds are derived from Integrated Public Use Microdata Series (IPUMS) microdata data that preserves and harmonizes decennial censuses from 1790 to 2010 and American Community Surveys (ACS). The dashed grey line plots the percentage of children ages 3 and 4 who are enrolled in full-year daycare. These data are as reported in the Current Population Survey (link: http://www.census.gov/hhes/school/data/cps/historical/.)

**Figure 2:**

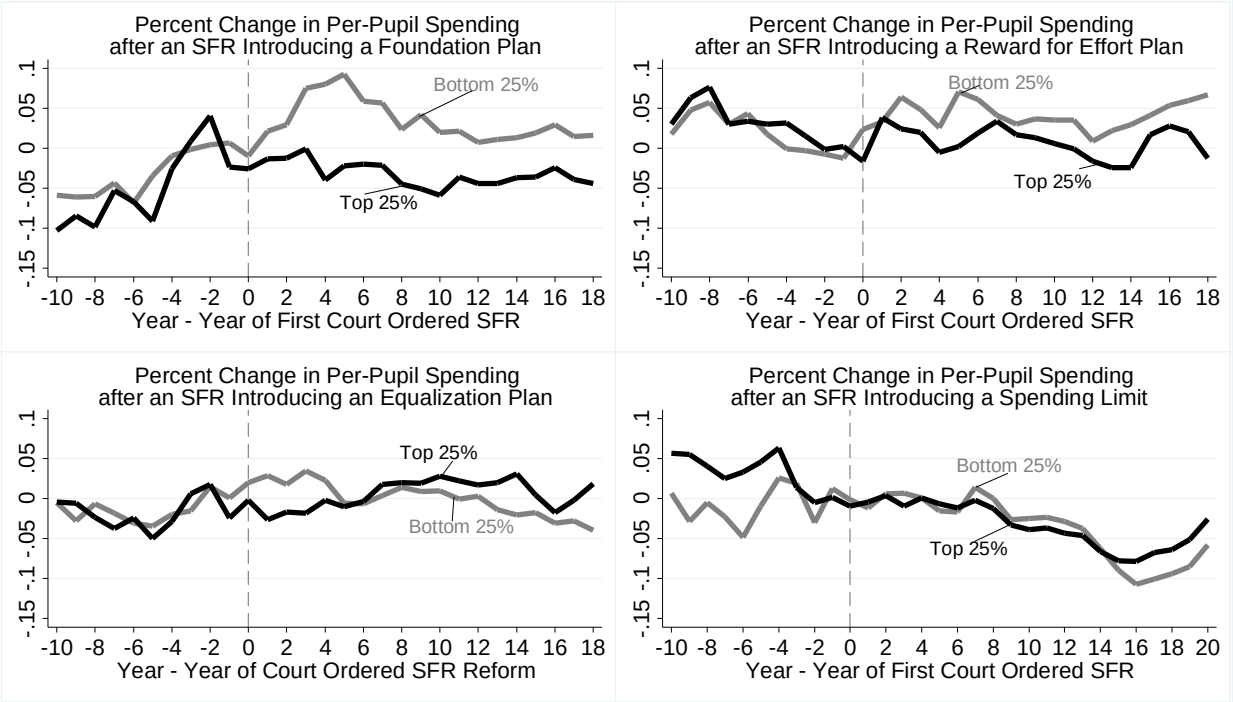
*Effect of Formula Type on District Per-Pupil Spending by District Spending in 1972*




Data: The sample includes all school districts in the United States between the years of 1967 and 2010. The sample is made up of 483,047 district-year observations. Each district is weighted by average enrollment for the full sample.
Model: These plots present the estimated event time coefficients of a regression on per-pupil spending at the district level on year fixed effects, district fixed effects, and the percentile group of the district in the state distribution of median income interacted with a full set of event-time indicator variables from 10 years prior to 19 years after the first court-mandated reform. The event-study plots are shown for the top and bottom 25 percent of districts in the state distribution of per-pupil spending in 1972. The event time plot has been re-centered at zero for the 10 pre-reform years so that the estimated coefficients represent the change in spending relative to the levels that persisted in the 10 years prior to the first reform.

**Figure 3:**

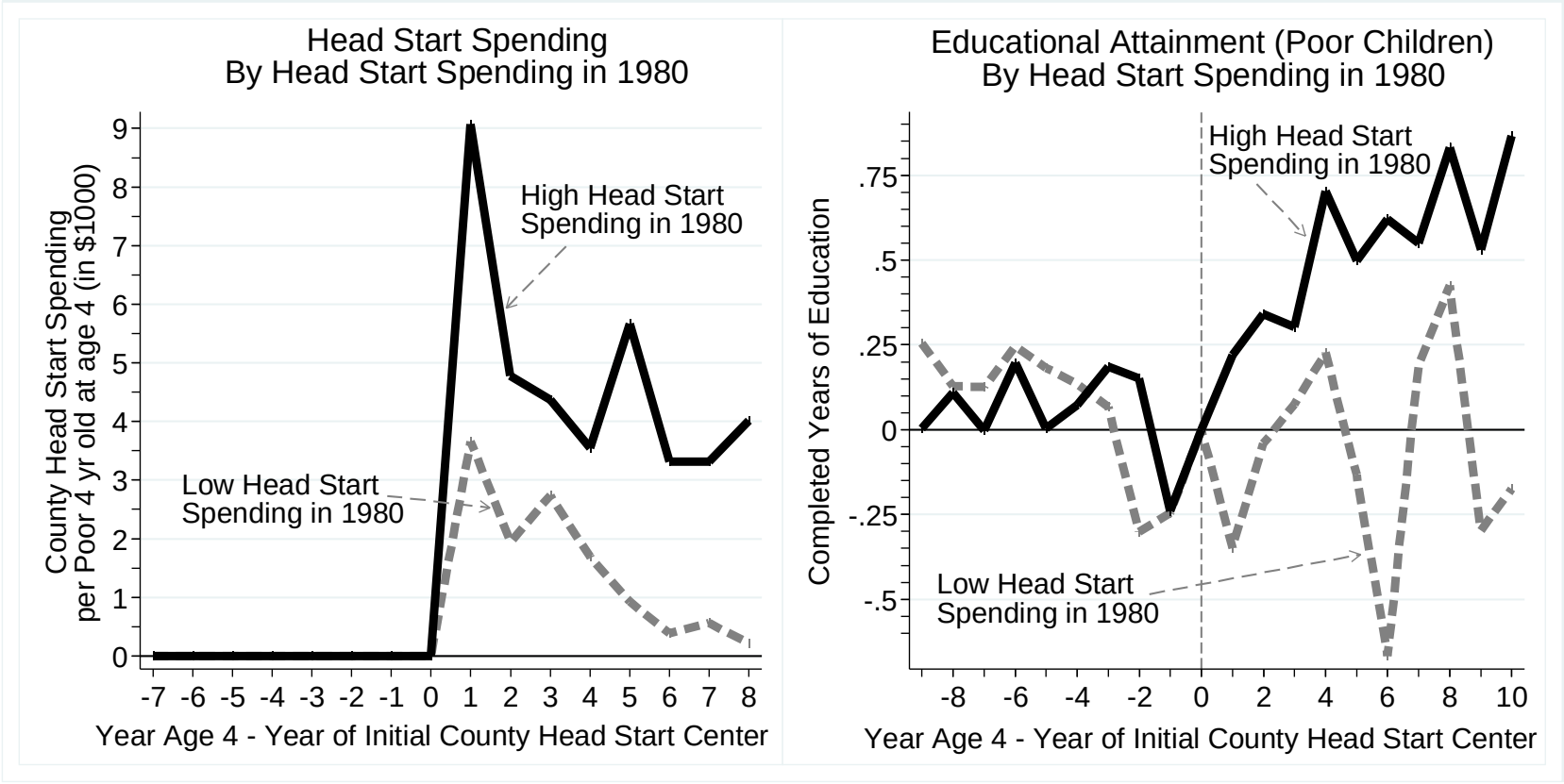
*Effect of Formula Type on District Per-pupil Spending by District Income in 1969*




Data: The sample includes all school districts in the United States between the years of 1967 and 2010. The sample is made up of 483,047 district-year observations. Each district is weighted by average enrollment for the full sample.
Model: These plots present the estimated event time coefficients of a regression on per-pupil spending at the district level on year fixed effects, district fixed effects, and the percentile group of the district in the state distribution of median income interacted with a full set of event-time indicator variables from 10 years prior to 19 years after the first court-mandated reform. The event-study plots are shown for the top and bottom 25 percent of districts in the state distribution of median family income in 1969. The event time plot has been re-centered at zero for the 10 pre-reform years so that the estimated coefficients represent the change in spending relative to the levels that persisted in the 10 years prior to the first reform.

**Figure 4:**

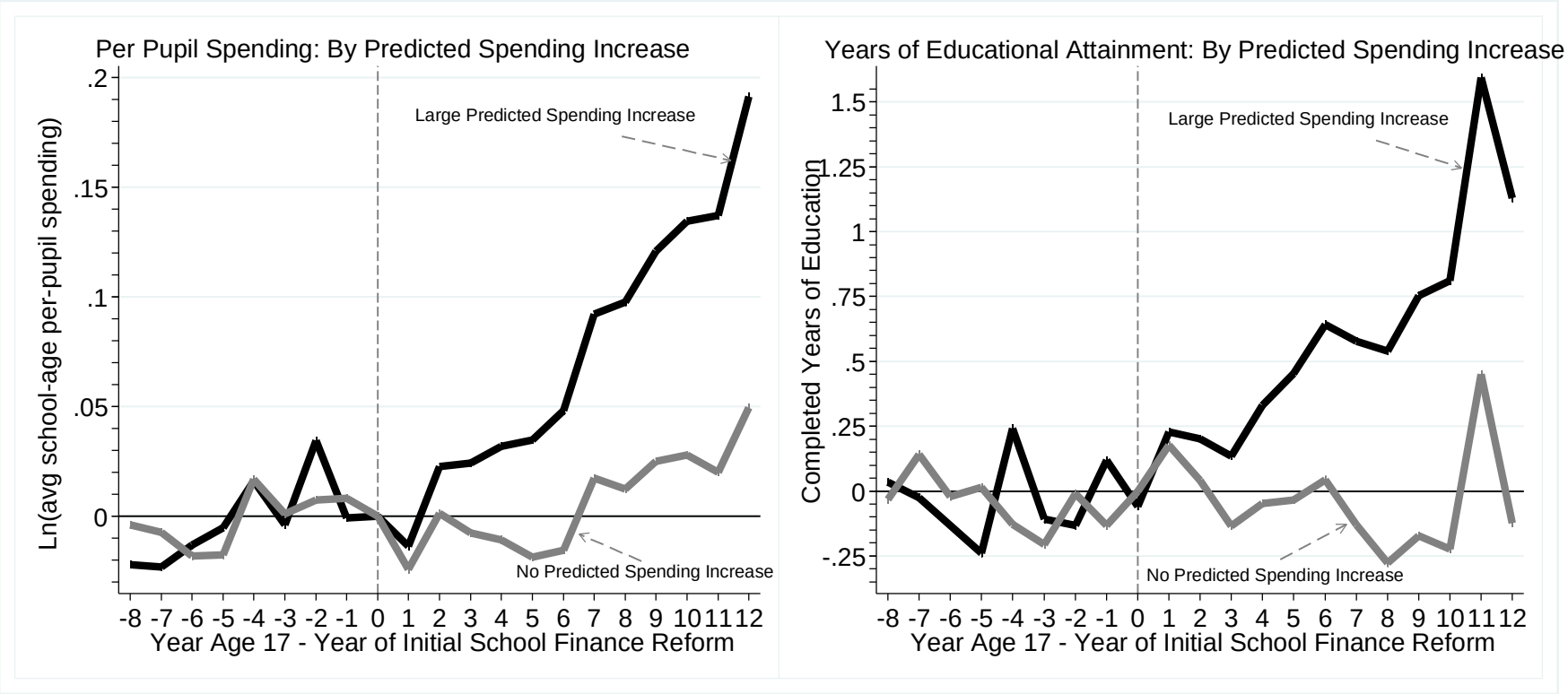
*Evolution of Head Start Spending and Educational Attainment at Rollout*




Data: Analysis sample includes PSID individuals born 1950-1976 who have been followed into adulthood. "High Head Start spending" is defined here as counties in the top quartile of Head Start spending among all US counties after rollout; "Low Head Start spending" defined here as bottom quartile of Head Start spending among all US counties after rollout or no spending.

Models: Results are based on event study models of educational attainment on children's exposure to county Head Start spending per poor 4-year-old at age 4 as a function of the timing of the rollout of the program in the county. The models include childhood county fixed effects, race*census division-specific birth year trends; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size) each interacted with linear cohort trends; and controls for childhood family characteristics (parental income & education, mother's marital status at birth, birth weight, gender). The figures present the results for low-income children.

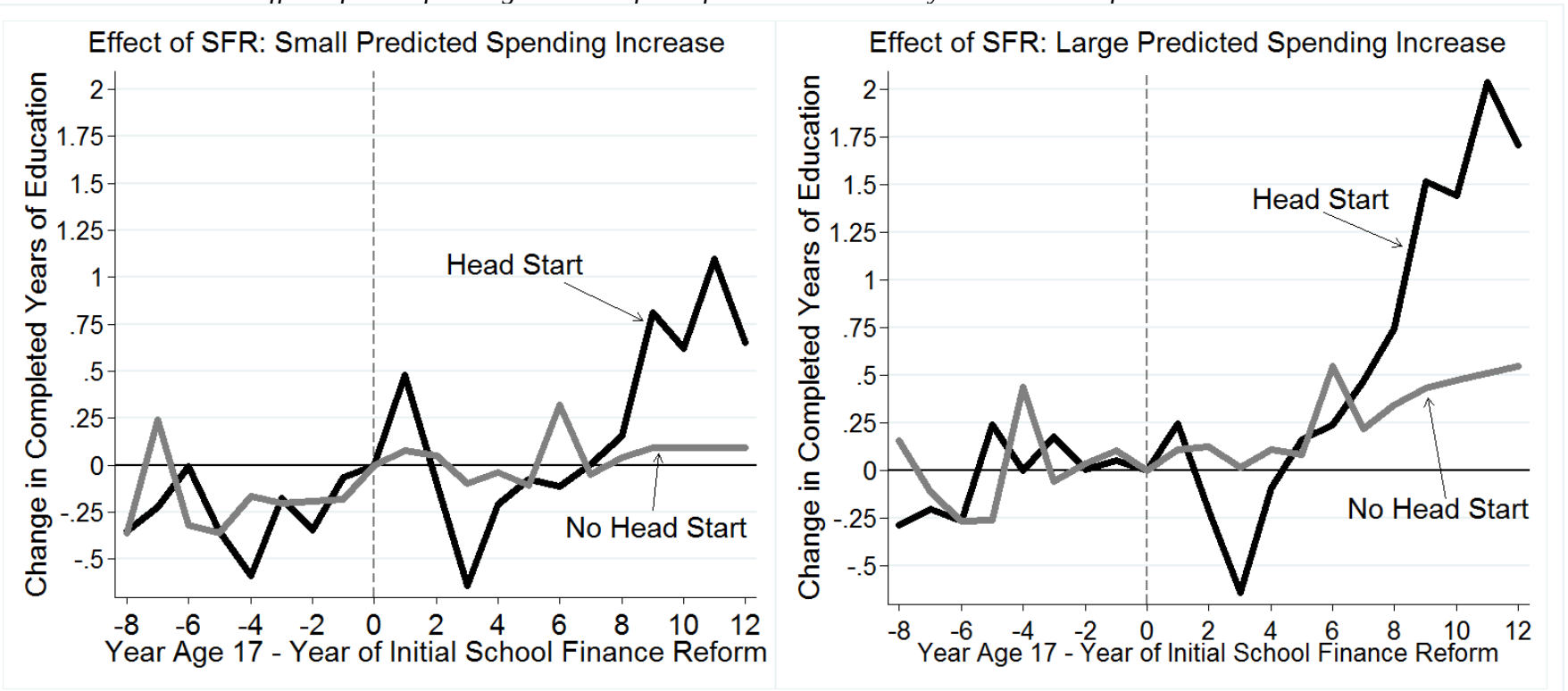
**Figure 5:**
*Evolution of K12 Spending and Educational Attainment after SFR Reform*




Models: The event study figures use school district's predicted reform-induced change in spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories—the solid black line shows estimated effects for districts with a predicted reform-induced K12 spending increase ($\widehat{dose}_d > 0$) whereas the solid grey line shows the corresponding effects for districts with low predicted reform-induced K12 spending increases or a decrease $\widehat{dose}_d \leq 0$. Roughly two-thirds of districts in reform states had predicted spending increases. The event study models include: school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, rollout of "War on Poverty" & related safety-net programs (community health centers, food stamps, medicaid, AFDC, UI, Title-I (average during childhood years)), timing of state-funded Kindergarten introduction and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size) each interacted with linear cohort trends; and controls for childhood family characteristics (parental income/education, mother's marital status at birth, birth weight, gender).

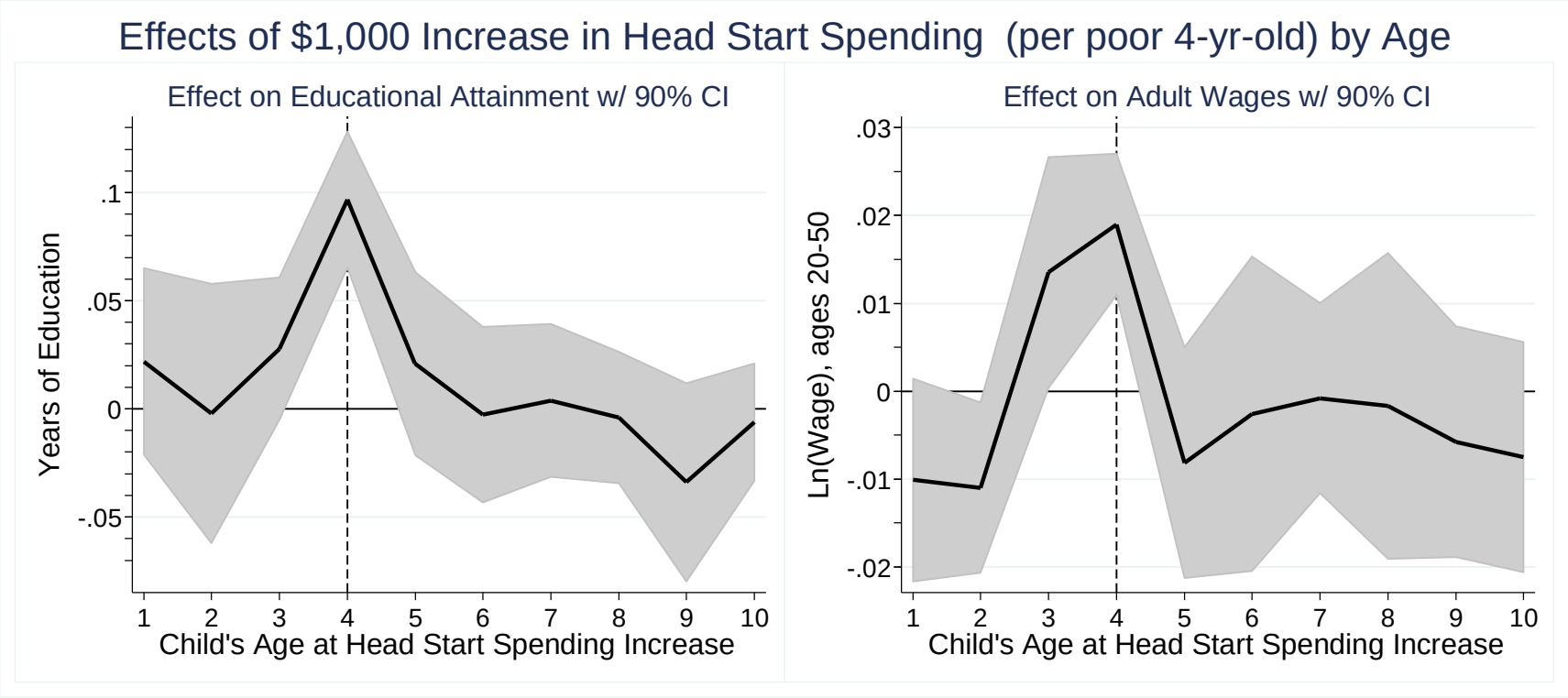
**Figure 6:**

*Effect of K12 Spending on Year of Completed Education: by Head Start Exposure Status*




Models: The event study figures use school district's predicted reform-induced change in spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories--right panel shows estimated effects for districts with a predicted reform-induced K12 spending increase ($\widehat{dose}_d > 0$) whereas the left panel shows the corresponding effects for districts with low predicted reform-induced K12 spending increases or a decrease $\widehat{dose}_d \leq 0$. Roughly two-thirds of districts in reform states had predicted spending increases. These estimated effects are presented both for children whose county had no Head Start center at age 4 (grey line), and those who were exposed to any county Head Start spending at age 4 (black line), to highlight the role of dynamic complementarity. The event study models include: school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, rollout of "War on Poverty" & related safety-net programs (community health centers, food stamps, medicaid, AFDC, UI, Title-I (average during childhood years)), timing of state-funded Kindergarten introduction and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size) each interacted with linear cohort trends; and controls for childhood family characteristics (parental income/education, mother's marital status at birth, birth weight, gender).

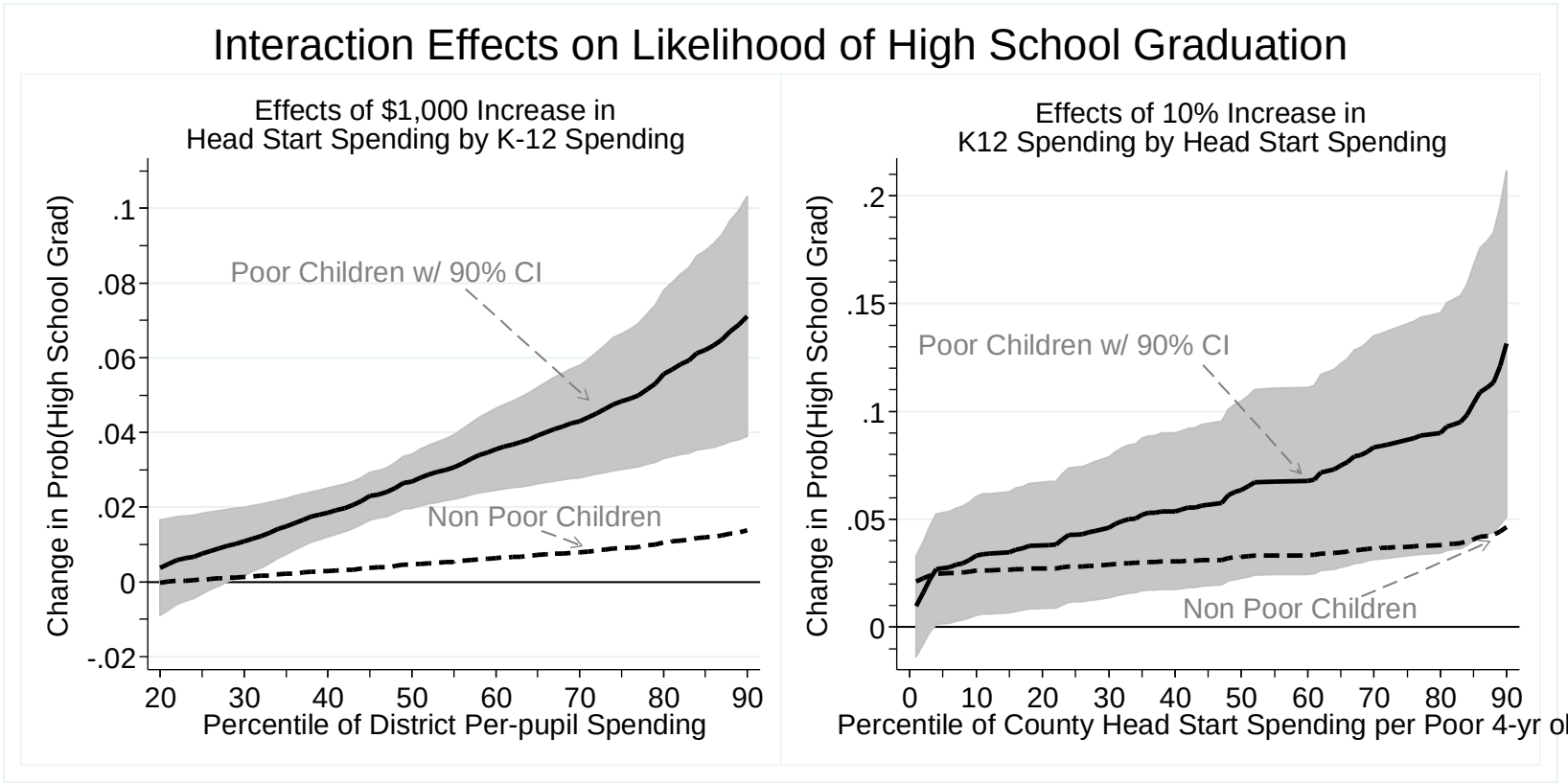
**Figure 7:**
*Effects of Head Start Spending: by Age of Spending*




These figures present the marginal effects of Head Start spending in an individual's childhood county at different ages, conditional on the level of Head start spending in the childhood county at age 4 (when such spending should have an effect). The sample is poor children only. Models include the full set of controls as in Tables 2 and 3. The coefficients on the non-eligible years 1 through 3 and 5 through 10, are all conditional on spending at age 4. The coefficient for spending at age 4 is based on model with no other ages included. For the regression estimates underlying this model for years of education attained, see Appendix Table G6.

**Figure 8:**

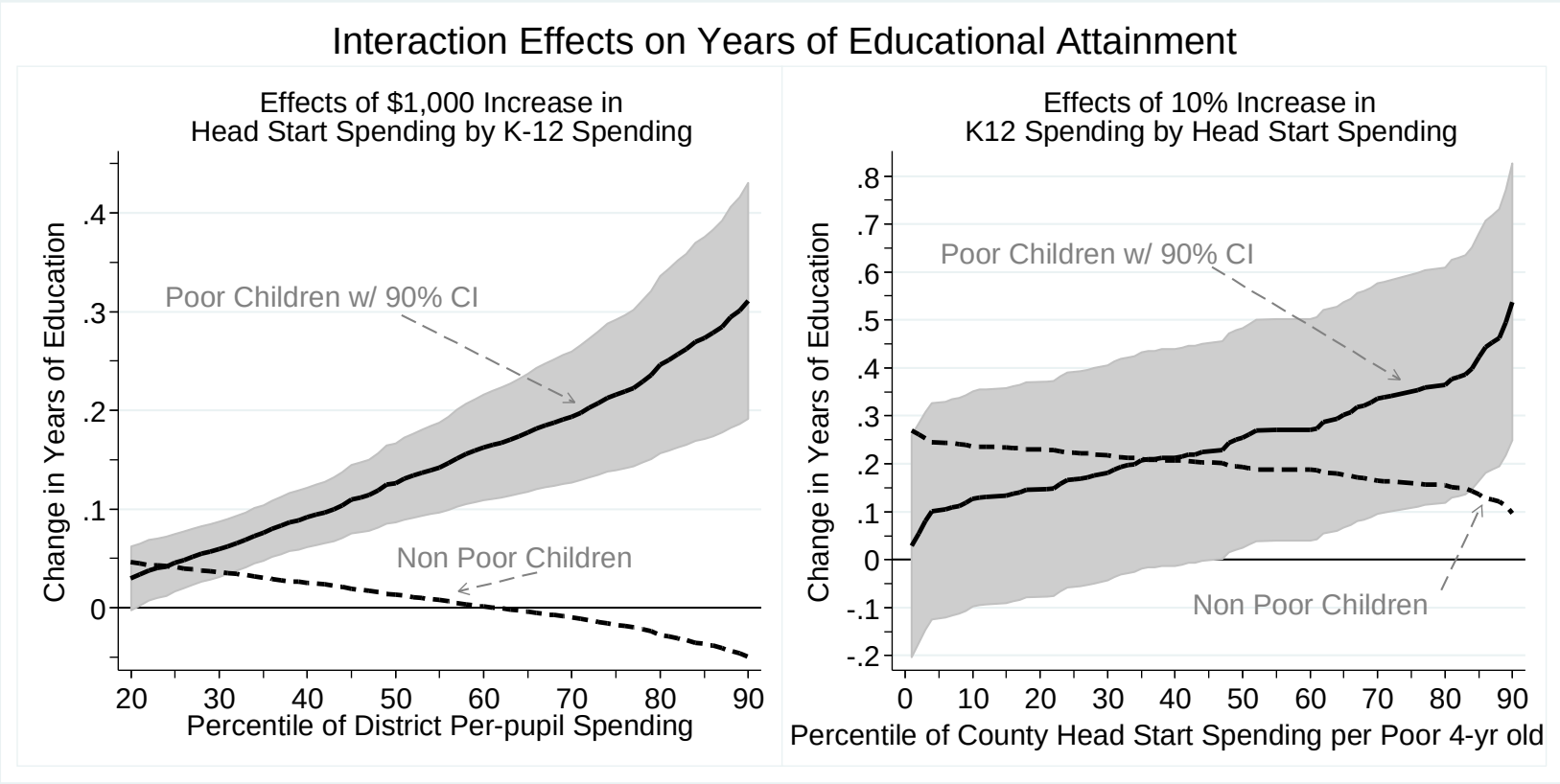
*Effect of Head Start Spending by K12 spending Levels and vice versa on High School Graduation*




Note: The reported marginal Effects based upon 2SLS-Difference-in-Difference model results are presented in columns 1 from Tables 2 and 3. The reported marginal effects and the standard errors were computed using the delta method.

**Figure 9:**

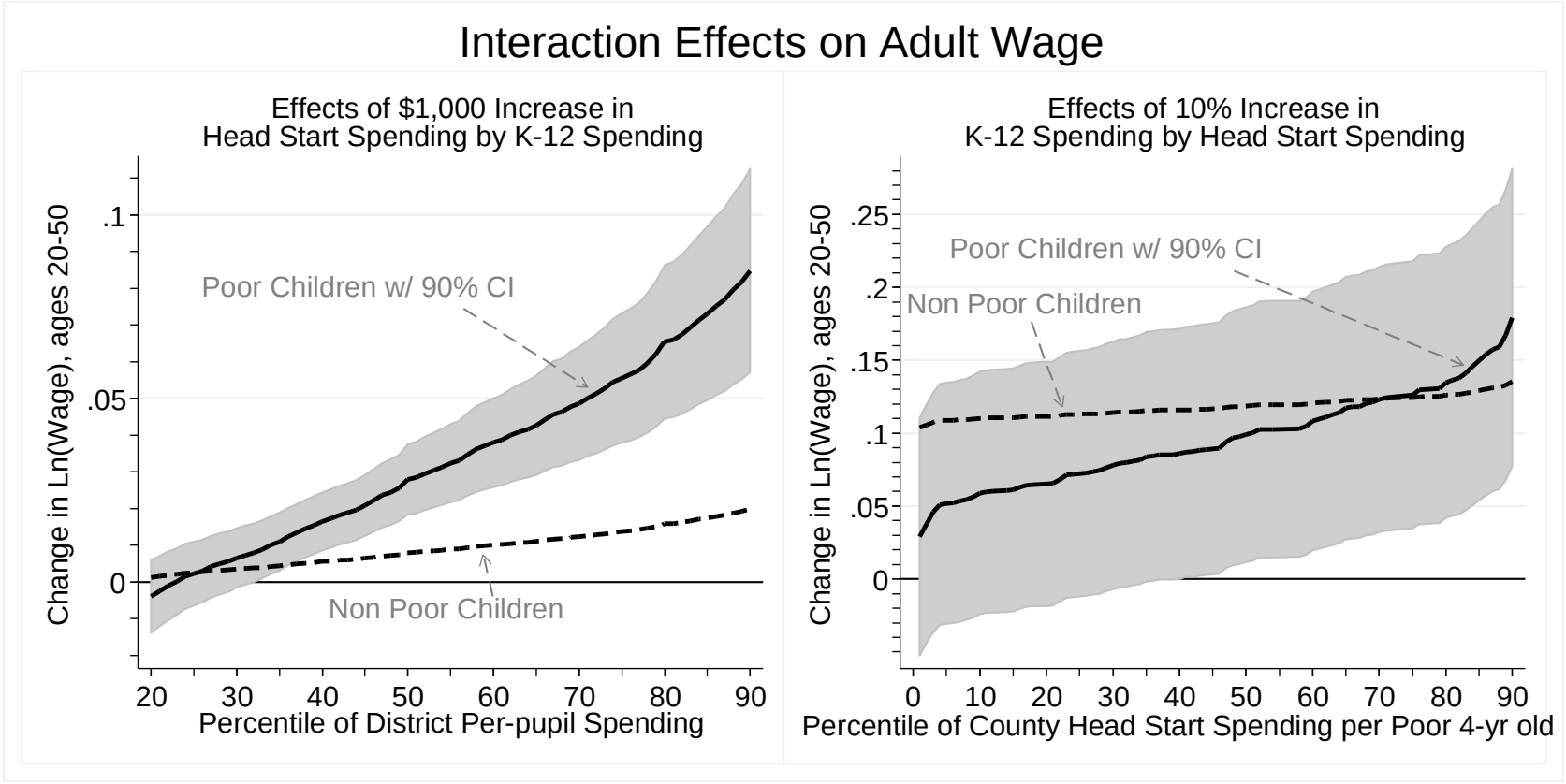
*Effect of Head Start Spending by K12 spending Levels and vice versa on Years of Education*




Note: The reported marginal Effects based upon 2SLS-Difference-in-Difference model results are presented in columns 2 from Tables 2 and 3. The reported marginal effects and the standard errors were computed using the delta method.

**Figure 10:**

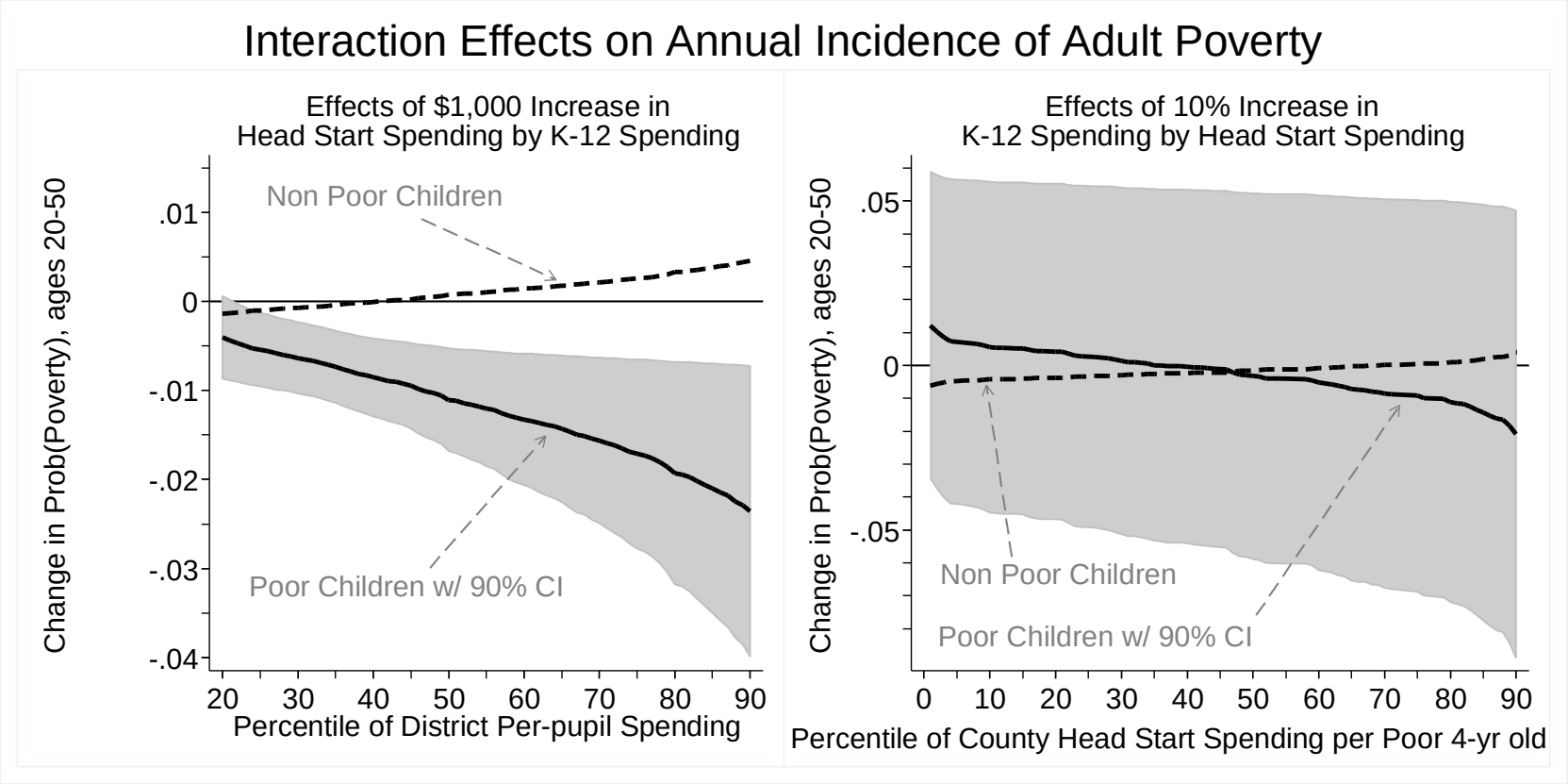
*Effect of Head Start Spending by K12 spending Levels and vice versa on Adult Wage (ages 20 -50)*




Note: The reported marginal Effects based upon 2SLS-Difference-in-Difference model results are presented in columns 3 from Tables 2 and 3. The reported marginal effects and the standard errors were computed using the delta method.

**Figure 11:**

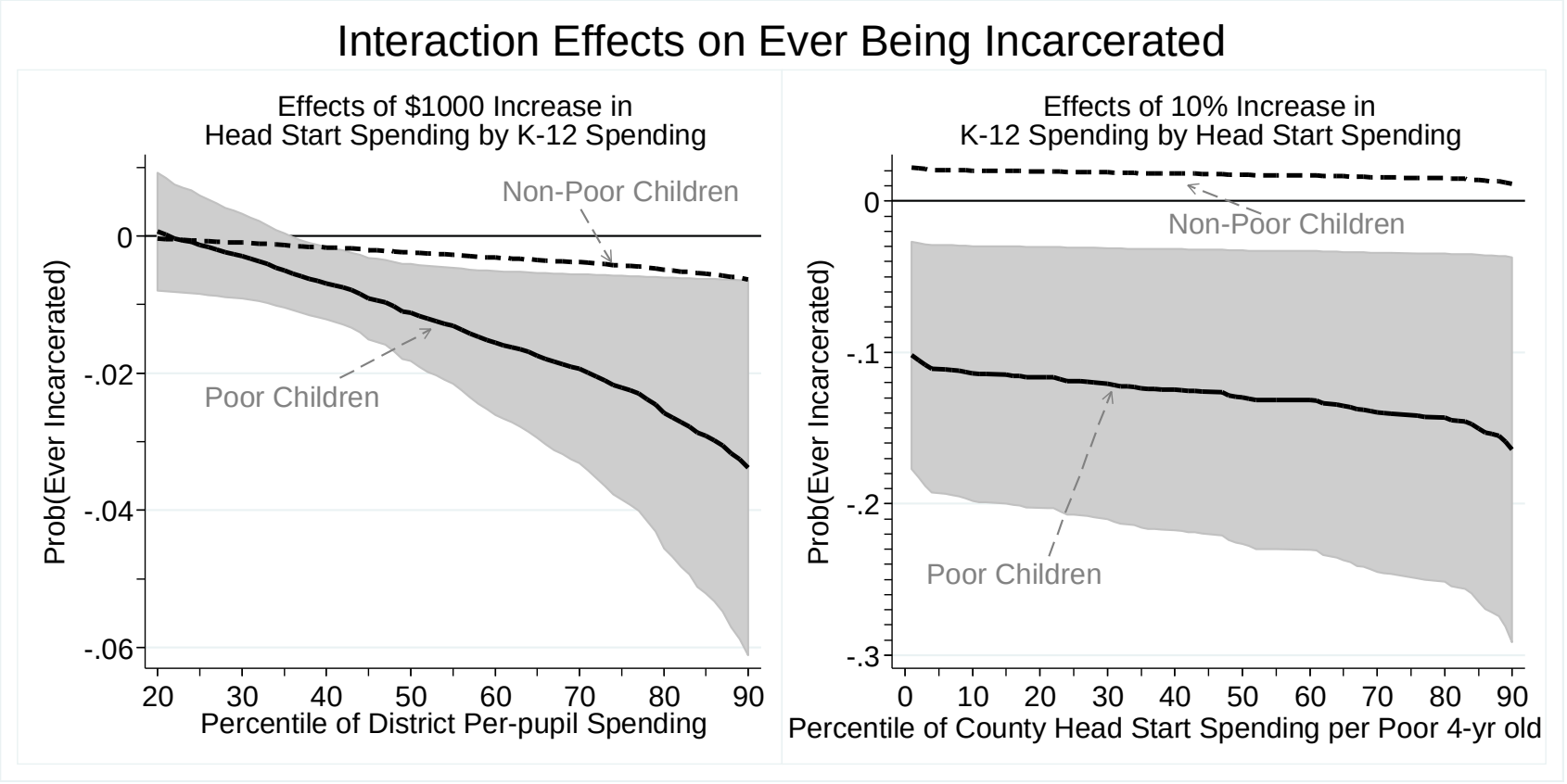
*Effect of Head Start Spending by K12 spending Levels and vice versa on Likelihood of Adult Poverty (ages 20 -50)*




Note: The reported marginal Effects based upon 2SLS-Difference-in-Difference model results are presented in columns 5 from Tables 2 and 3. The reported marginal effects and the standard errors were computed using the delta method.

**Figure 12:**

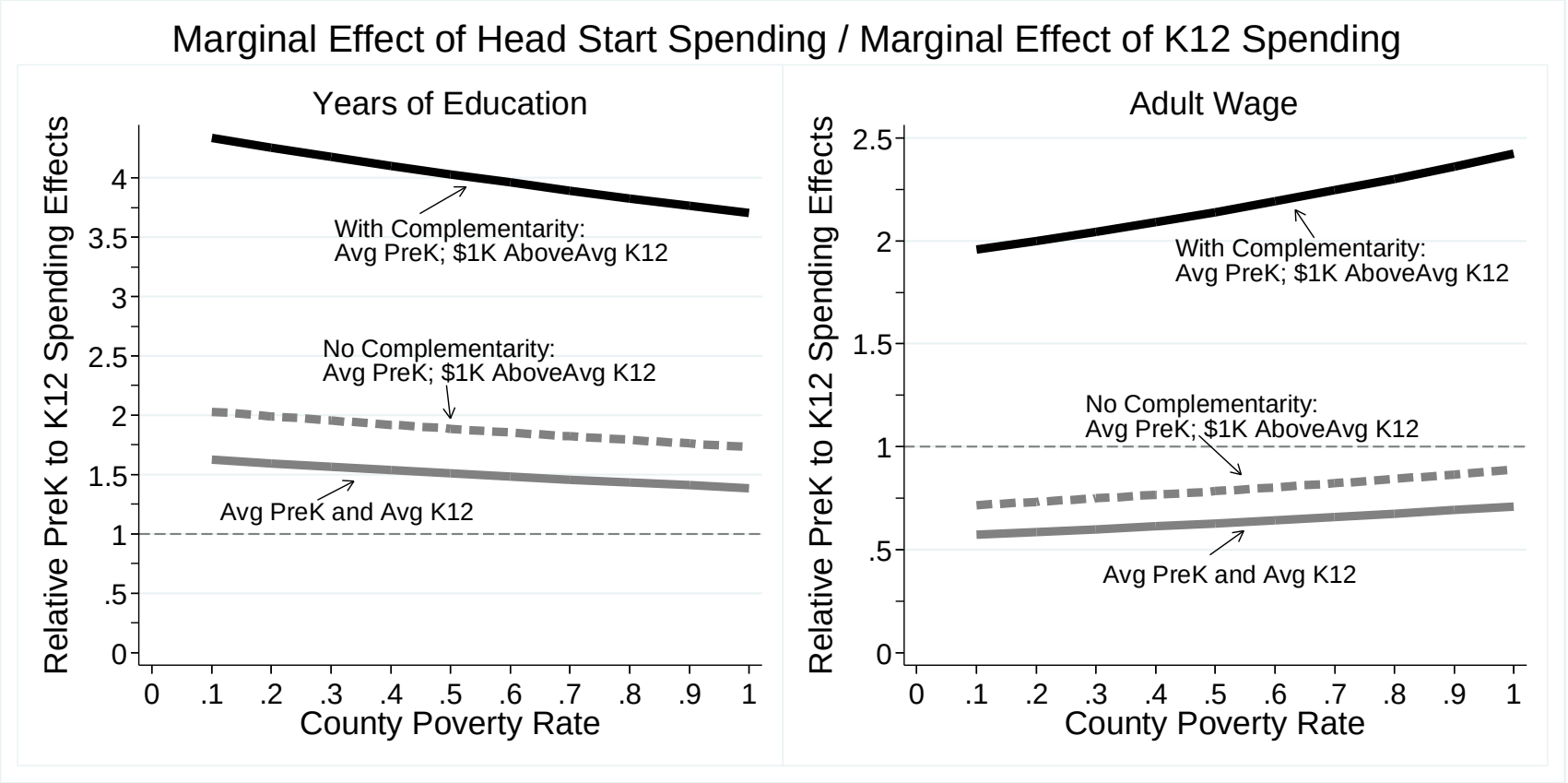
*Effect of Head Start Spending by K12 spending Levels and vice versa on ever Being Incarcerated*




Note: The reported marginal Effects based upon 2SLS-Difference-in-Difference model results are presented in columns 6 from Tables 2 and 3. The reported marginal effects and the standard errors were computed using the delta method.

**Figure 13:**

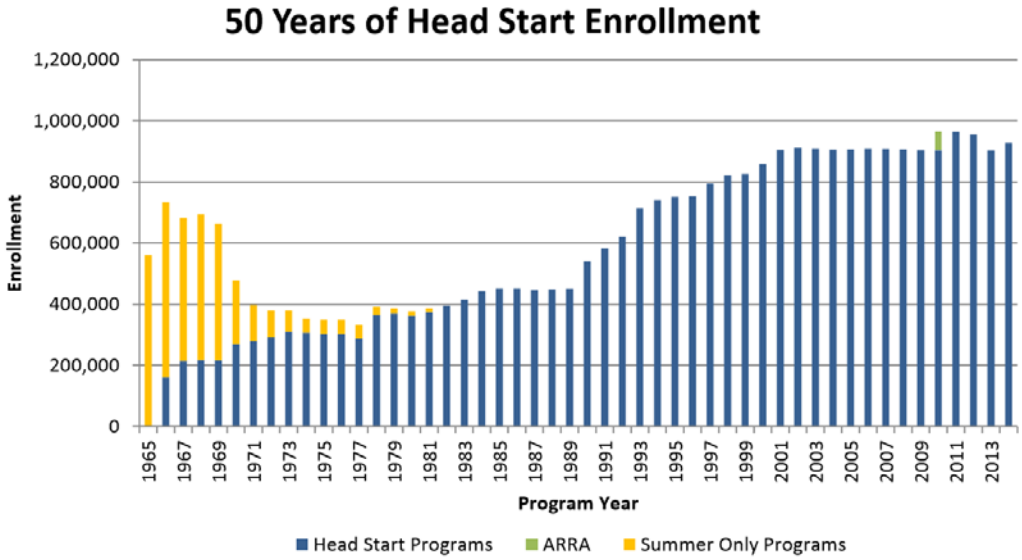
*Ratio between the Effect of Head Start Spending and K12 spending Levels by Poverty Level in the County*




Note: The reported marginal Effects based upon 2SLS-Difference-in-Difference model results presented in columns 2 and 3 from Tables 2 and 3. The solid grey lines plot the ratio between the marginal effect of spending on Head Start and the effect of spending that same amount on the K12 system (in present value-terms). This ratio presented in the solid grey line is evaluated at average levels of Head Start spending and K12 spending during the sample period. The dashed grey line presents this same ratio evaluated at $1000 above the average K12 spending levels assuming no dynamic complementarity, while the solid black line presents this ratio evaluated at $1000 above the average K12 spending levels using the estimated interaction effects. The difference between the solid black lines and the dashed grey lines reflect the marginal contribution of dynamic complementarity to changes in this ratio as one increases K12 spending above the average.

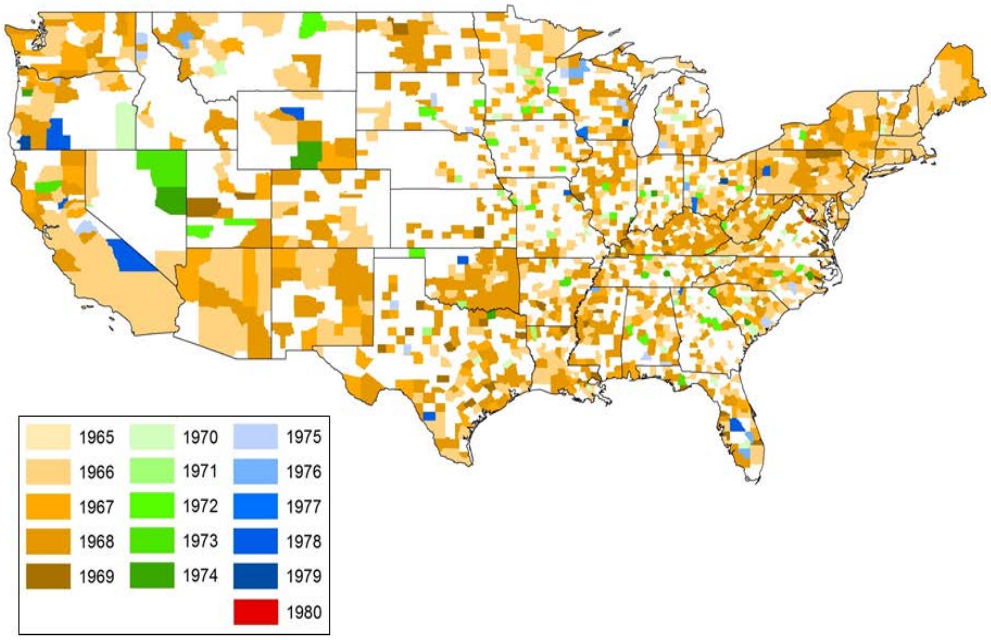
## Appendix A
*Additional Tables*

**Figure A1:**
*National Head Start Enrollment over Time*




Note: Chart is pasted directly from the Head Start Fact sheet (link: https://eclkc.ohs.acf.hhs.gov/hslc/data/factsheets/2015-hs-program-factsheet.html)

**Figure A2:**
*Year of Establishment of First Head Start Center by County*




Note: Based on authors' calculations and data collections as described in Appendix D.

## Appendix B

*Spending per enrollee versus spending per eligible*

**Table B2:**

*Relationship between Spending per Enrollee, Spending per Poor 4-Year-Old and Enrollment (at state-year level)*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| | Spending per Enrollee | | Log of Spending per Enrollee | | Log of Head Start Enrollment | | Share of Income Eligible four-year-olds enrolled | |
| Spending per poor 4-year-old | 0.0174 | 0.0379* | | | | | | |
| | [0.0359] | [0.0143] | | | | | | |
| Log Spending per poor 4-year-old | | | -0.0192 | 0.0243* | 0.0810+ | 0.121** | 0.648** | |
| | | | [0.0271] | [0.00930] | [0.0482] | [0.0438] | [0.0506] | |
| Log Spending per Enrollee | | | | | | | | -0.13 |
| | | | | | | | | [0.208] |
| Year FX | N | Y | N | Y | N | Y | Y | Y |
| State FX | Y | Y | Y | Y | Y | Y | Y | Y |
| Observations | 612 | 612 | 612 | 612 | 612 | 612 | 612 | 612 |
| R-squared | 0.759 | 0.927 | 0.79 | 0.93 | 0.996 | 0.996 | 0.984 | 0.909 |

Robust standard errors in brackets adjusted for clustering at the state level

** p<0.01, * p<0.05, + p<0.1

Notes: State year level data on total federal Head Start spending and total Head Start enrollment is obtained from the Head Start Facts fiscal years reports 1999 through 2015. Data on the number of poor four-year-olds in the state in each year is obtained from Integrated Public Use Microdata Series (IPUMS) microdata that preserves and harmonizes decennial censuses from 1790 to 2010 and American Community Surveys (ACS).

## Appendix C:

*Matching PSID Individuals to their Childhood School Districts*

In order to limit the possibility that school district boundaries were drawn in response to pressure for SFRs, we utilize 1969 school district geographies. The "69-70 School District Geographic Reference File" (Bureau of Census, 1970) relates census tract and school district geographies. For each census tract in the country, it provides the fraction of the population that is in each school district. Using this information, we aggregate census tracts to 1970 district geographies with Geographic Information Systems (GIS) software. We assign census tracts from 1960, 1980 and 1990 to school districts using this resulting digital map based on their centroid locations.

To construct demographic information on 1969-1970-definition school districts, we compile census data from the tract, place, school district and county levels of aggregation for 1960, 1970, 1980 and 1990. We construct digital (GIS) maps of 1970 geography school districts using the 1969-1970 School District Geographic Reference File from the Census. This file indicates the fraction by population of each census tract that fell in each school district in the country. Those tracts split across school districts we allocated to the school district comprising the largest fraction of the tract's population. Using the resulting 1970 central school district digital maps, we allocate tracts in 1960, 1980 and 1990 to central school districts or suburbs based on the locations of their centroids. The 1970 definition central districts located in regions not tracted in 1970 all coincide with county geography which we use instead.

## Appendix D

### *County-Level Federal Outlays for Head Start and Title I, 1965-1980*

Our collection of head Start data follows Johnson (2015). County-year federal outlays for Head Start and Title I ESEA were computed using county-level federal outlays data acquired from the National Archives and Records Administration (NARA) for fiscal years 1965 through 1980, along with ICPSR Study #6029 (for fiscal years 1976 to 1980). Information was culled from NARA records by searching program titles and program codes. We identified the pool of grants for Head Start from the NARA records, which included string searches on Head Start grant titles. For most records, Head Start programs are listed by community and funding amounts, and information on the "stock" of programs at a particular time allows verification of the accuracy of grant "flows". Likewise, we identified the pool of grants for Title I/ESEA outlays from the NARA records by using program titles and program codes over this period. The county-year federal Head Start and Title I outlays were converted into 2000 dollars using the CPI-U deflator.

County-level information on Community Action Program (CAP) Grants and grantees on federal CAP grants is derived from the NARA microdata (Community Services Administration 1981). These data files document neighborhood and community based poverty programs as funded by CAP and CAP grant-action data include data on the target population of grant proposals. These records are structured as two data files spanning 1965 through 1980. One data set is observed at the level of individual grant actions; the other data set records data on the organizations receiving grants. The combined data include information on any “action” on a grant (when it is recorded, extended, renewed, or terminated), dates associated with these actions, and some information about the funded project. We use the county-level geographical identifiers from the grantee data and grant-action file, which include the name and county of designated grantee and county where the services are provided in most cases. We aggregate these amounts by the fiscal year of disbursement and county of service delivery. These amounts have been verified by state against information printed in OEO annual reports (Office of Economic Opportunity, 1965–1968).

We compared our calculated county-level federal outlays for Head Start with those reported in Ludwig and Miller (2007) for fiscal years 1968 and 1972, and Elizabeth Cascio (2009) for 1976-80, and in each case our numbers line up with those used by these authors (who generously shared their data for comparison). Our county-level panel of Head Start spending though spans a much longer time period than used in previous studies. We compared spending totals calculated from the county-level files to published data at the federal level and state level (where available) to assess the validity of the county-level data. Following Cascio (2009), we compared the state-level Head Start outlays calculated in our data to those reported in Jones (1979) for fiscal years 1970 through 1977, and the correlation coefficient was above 0.975 in all fiscal years except 1974, where Mississippi was an obvious outlier. We therefore dropped all fiscal years for Mississippi for the Head Start analysis because the reporting of federal outlays for that state at the county-level had some obvious errors and were poorly documented.

We then assembled population counts of the number of 4-year olds and the number of school-age children ages 5-17 in every US county, respectively, using the Surveillance, Epidemiology, & End

Results (SEERS) program data spanning the period 1965 through 1980. The county-year federal outlays for Head Start and Title I ESEA were combined with both the county-year population counts of the number of 4-year olds and number of children ages 5-17, and the 1970 county-level poverty rates among children (and non-elderly persons) in order to construct our measures of county-level Head Start spending per poor 4-year-old and county-level Title I (ESEA) spending per-pupil, for 1965 through 1980.[37]

***District-Level K-12 School Spending Data***

Previous historical data on per-pupil expenditures was only available in a readily usable format via the *Census of Governments: School System Finance (F-33) File* (U.S. Bureau of the Census, Department of Commerce). The Census of Governments previously was only conducted in years that end in a two or seven, so at the time when many important papers on SFRs were written, there were many years of missing data. In addition, until recently the earliest available F-33 data was for the year 1972. As a result, it was previously impossible to model per-pupil spending and spending inequality annually over time, so many authors (e.g., MES, Card and Payne), operating under the Common Trends Assumption, assumed that trends in per-pupil spending were linear. Due to these limitations, previous papers on school finance reforms were also unable to look at how the exact timing of reforms affected per-pupil expenditure and spending inequality within a state.

Our data from the Historical Database on Individual Government Finances (INDFIN) represents the Census Bureau's first effort to provide a time series of historically consistent data on the finances of individual governments. This database combines data from the *Census of Governments Survey of Government Finances (F-33)*, the National Archives, and the *Individual Government Finances Survey*. The School District Finance Data FY 1967-91 is available annually from 1967 through 1991. It contains over one million individual local government records, including counties, cities, townships, special districts, and independent school districts. The INDFIN database frees the researcher from the arduous task of reconciling the many technical, classification, and other data-related changes that have occurred over the last 30 years. For example, this database includes corrected statistical weights that have been standardized across years, which had not been done previously. Furthermore, although most governments retain the ID number they are assigned originally, there are circumstances that result in a government's ID being changed. Since a major purpose of the INDFIN database is tracking government finances

[37] References for the data appendix:


- Cascio, Elizabeth (2009). "Do Investments in Universal Early Education Pay Off? Long-term Effects of Introducing Kindergartens into Public Schools". NBER Working Paper No. 14951.
- Johnson, Rucker C. (2015). "Follow the Money: School Spending from Title I to Adult Earnings". Special edited volume, ESEA at 50, published in *The Russell Sage Foundation Journal of the Social Sciences*.
- Jones, Jean Yavis (1979). "The Head Start Program – History, Legislation, Issues and Funding 1964-1978". Washington, D.C.: Congressional Research Service. Report 79-14 EPW.
- Ludwig, Jens and Douglas L. Miller (2007). "Does Head Start Improve a Children's Life Chances? Evidence from a Regression Discontinuity Design." *Quarterly Journal of Economics* 122(1): 159-208.
- Office of Economic Opportunity. Annual Reports. Washington, DC: GPO, 1965–1968.
- U.S. Bureau of the Census. Census of Population Supplementary Report. Poverty Status in 1969 and 1959 of Persons and Families, for States, SMSA's, Central Cities and Counties: 1970 and 1960.

over time, it is critical that a government possess the same ID for all years (unless the ID change had a major structural cause). For example, All Alaska IDs were changed in the 1982 Census of Governments. In addition, new county incorporations, where governments in the new county area are re-assigned an ID based on the new county code (e.g., La Paz County, AZ), cause ID changes. Thus, if a government ID number was changed, the ID used in the database is its current GID number, including those preceding the cause of the change, so that the ID is standardized across years.

In addition to standardizing the data, the Census Bureau has corrected a number of errors in the INDFIN database that were previously in other sources of data. For example, for fiscal years 1974, 1975, 1976 and 1978 the school district enrollment data that had previously been released were useless (either missing or in error for many records). Thus, in August 2000, these missing enrollment data were replaced with those from the employment survey individual unit files. This enables us to more accurately compute per-pupil expenditures for those years. In addition, source files before fiscal 1977 were in whole dollars rather than thousands. This set a limit on the largest value any field could hold. If a figure exceeded that amount, then the field contained a special "overflow" flag (999999999). Few governments exceeded the limit (Port Authority of NY and NJ and Los Angeles County, CA are two that did). For the INDFIN database, actual data were substituted for the overflow flag. Finally, in some cases the Census revised the original data in source files for the INDFIN database. In some cases, official revisions were never applied to the data files. Others resulted from the different environment and operating practices under which source files were created. Finally, some extreme outliers were identified and corrected (e.g., a keying error for a small government that ballooned its data).

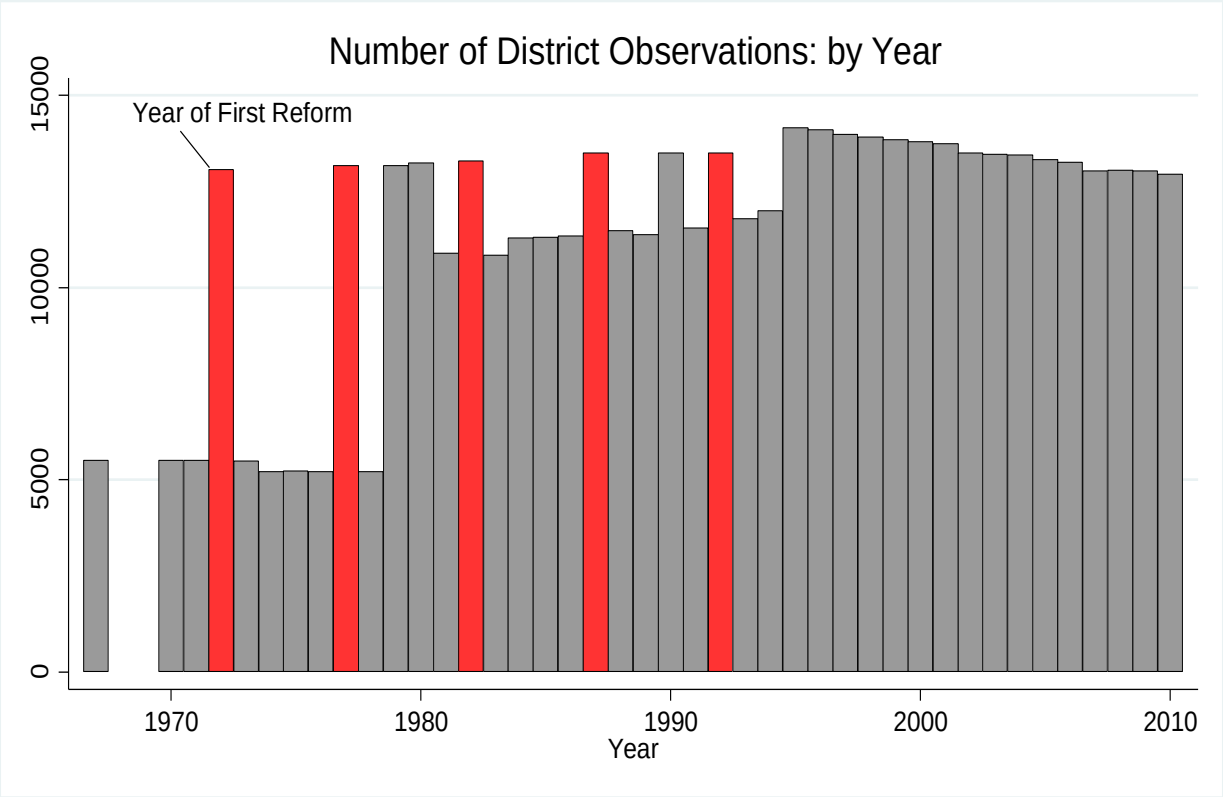
The Common Core of Data (CCD) School District Finance Survey (F-33) consists of data submitted annually to the National Center for Education Statistics (NCES) by state education agencies (SEAs) in the 50 states and the District of Columbia. The purpose of the survey is to provide finance data for all local education agencies (LEAs) that provide free public elementary and secondary education in the United States. Both NCES and the Governments Division of the U.S. Census Bureau collect public school system finance data, and they collaborate in their efforts to gather these data. The Census of Governments, which was recorded every five years until 1992, records administrative data on school spending for every district in the United States. After 1992, the Public Elementary-Secondary Education Finances data were recorded annually with data available until 2010. We combine these data sources to construct a long panel of annual per-pupil spending for each school district in the United States between 1967 and 2010.

Per-pupil spending data from before 1992 is missing for Alaska, Hawaii, Maryland, North Carolina, Virginia, and Washington, D.C. Per-pupil spending data from 1968 and 1969 is missing for all states. Spending data in Florida was also missing for 1975, 1983, 1985-1987, and 1991. Spending data in Kansas was also missing for 1977 and 1986. Spending data in Mississippi was also missing for 1985 and 1988. Spending data in Wyoming was also missing for 1979 and 1984. Spending data for Montana is missing in 1976, data for Nebraska is missing in 1977, and data for Texas is missing in 1991. Where there was only a year or two of missing per-pupil expenditure data, we filled in this data using linear interpolation.

Figure D1 below shows the number of district observations in our data for each year. The bars highlighted in red are the census of government years employed in previous national studies of school finance reforms (e.g. Card and Payne 2002, Hoxby 2001, Murray Evans and Schwab 1998). While the coverage of the data we use is arguably better that that used previously, it is not perfect. As shown in Appendix figure D1, for years, 1967, 1970, 1971, 1973, 1974, 1975, 1976,

and 1978 only about 40 percent of districts are present (often larger districts). After 1979 almost all districts are included.

**Figure D1:** *The number of district observations for each year.*




***Data on School Finance Reforms***

Due to great interest on the topic, the timing of school finance reforms (SFRs) has been collected in various places. Data on the exact timing and type of court-ordered and legislative SFRs was obtained from Public School Finance Programs of the Unites States and Canada (PSFP), National Access Network's state by state school finance litigation map (2011), from Murray, Evans, and Schwab (1998), Hoxby (2001), Card and Payne (2002), Hightower et al (2010), and Baicker and Gordon (2004). The most accurate information on school finance laws can be derived from the PSFP, which provides basic information and references to the legislation and court cases challenging them (Hoxby 2001). In most cases, data from these sources are consistent with each other. Where there are discrepancies we often defer to PSFP, but also consulted LexisNexis and state court and legislation records.

There were discrepancies in reported timing of overturned court cases in several states: Connecticut (Hoxby states the decision was made in 1978, but Card and Payne report it was made in 1977), Kansas (Hoxby states 1976, but PSFP and ACCESS report 1972), New Jersey (Card and Payne state 1989, but PSFP says 1990), Washington (Murray, Evans, and Schwab, Hoxby, and Card and Payne report 1978, but PSFP reports 1977), Wyoming (Hoxby says 1983, but Card and

Payne and Murray, Evans, and Schwab report 1980). We researched each case by name to discover the true date of the decision.

Using a policy survey conducted during the 2008-2009 school year, a recent study by Hightower et al (2010) provides a description of state finance policies and practices. This study was used to verify whether there had been any changes to state funding formulas between 1998 and 2009. We only collected information on the first five court cases per state in which the state found the school funding system unconstitutional. There were only three states with five or more court cases overruling the funding system (New Hampshire, New Jersey, and Texas). In addition, we only collected information on the first four court cases per state in which states upheld the school funding system. There were only four states with four or more court cases in which the school funding system was upheld (Illinois, New York, Oregon, and Pennsylvania).

Information on whether or not a state funding formula had a MFP, flat grant formula, variable matching grant scheme, recapture provision, spending limit, power equalization scheme, local-effort equalization scheme, or full state funding came from *PSFP* (1998) and was verified using Card and Payne (2002) and Hightower et al (2010). We defined MFPs, flat grant formulas, and variable matching grant schemes in the same way as Card and Payne did in their 2002 study. We defined power equalization, local-effort equalization, and full state funding in the same way as the EPE study (Hightower, Mitani and Swanson 2010). Each element of a state funding formula was coded as a dichotomous variable. For example, MFP is a dichotomous variable that is equal to one in the year and all subsequent years in which a state's finance system had a MFP plan in place. MFP was set equal to zero in all years prior to the state's funding system having a MFP in place, or if a state never implemented a MFP. Information on the timing of spending and tax limits came from Downes and Figlio (1998). We also supplemented this with data from *PSFP* for years after those covered in Downes and Figlio (1998).

## Appendix E
*Predicting Dosage*

The prediction of $dose_c$ is obtained in two steps. First, using district-by-birth-cohort data for the full universe of districts (not only those represented in the PSID), we estimate [E1] where all variables are defined as in [5], $I_{F,d}$ is an indicator for the type of reform (*F*) (i.e– *foundation plans, spending limits, reward for effort plans, equalization plans, and equity cases*) introduced by the court order in the state containing district *d*, $Q_{ppe72,d}$ is the quartile of district *d* in the state distribution of per-pupil spending in 1972, and $Q_{inc69,d}$ is the quartile of district *d* in the state distribution of median income in 1969.

[E1] $$\ln(PPE_{5-17})_{idb} = \sum_{Q_{ppe=1}}^{4}\sum_{T=-20}^{20}\left(I_{T_{idb}=T}\times I_{Q_{ppe72,d}=Q_{ppe}}\right)\cdot\alpha_{T,Q_{ppe}} + \sum_{F=1}^{5}\sum_{Q_{ppe=1}}^{4}\sum_{T=-20}^{20}\left(I_{T_{idb}=T}\times I_{Q_{ppe72,d}=Q_{ppe}}\times I_{F,d}\right)\cdot\alpha_{T,Q_{inc},F} + \Pi C_{idb} + \theta_{d3} + \theta_{b3} + \varphi_{idb}.$$

The coefficients $\alpha_{T,Q_{ppe}}$ map out the effect of *T* years of exposure to a court-ordered SFR for those from districts in the $Q^{th}$ quartile of the state distribution of per-pupil spending in 1972. Similarly, the coefficients $\alpha_{T,Q_{inc},F}$ map out the effects on school-age per-pupil spending of *T* years of exposure to a court-ordered SFR that introduced reform type *F* for those from districts in the $Q^{th}$ quartile of the state distribution of median income in 1969. Second, for each district we use the predicted spending change (based on reform type implemented by the state and district spending and district income levels prior to reforms) for those who were between the ages of 10 and 15 in the year of the initial court-ordered SFR. As such, our predicted effect from [E1] is

[E2] $$\widehat{\text{dose}}_d = \left[\sum_{Q_{ppe=1}}^{4}\sum_{T=-20}^{20}\left(I_{T_{idb}=T}\times I_{Q_{ppe72,d}=Q_{ppe}}\right)\cdot\hat{\alpha}_{T,Q_{ppe}} + \sum_{F=1}^{5}\sum_{Q_{ppe=1}}^{4}\sum_{T=-20}^{20}\left(I_{T_{idb}=T}\times I_{Q_{ppe72,d}=Q_{ppe}}\times I_{F,d}\right)\cdot\hat{\alpha}_{T,Q_{inc},F}\right]/6$$

By using the predicted values, $\widehat{\text{dose}}_d$, from [E2] from the full universe of school districts as an instrument in a 2SLS regression on the PSID sample, we implement a two-sample instrumental variables (2S-2SLS) strategy where our excluded instruments are the exposure indicator variables interacted with a function of the reform type implemented by the state, the district income level prior to reforms, and the spending level of the district prior to reforms.[38] This approach captures meaningful variation in K12 spending due to the reforms but removes any variation in spending that is determined by local factors that also influence outcomes.

[38] The two-sample 2SLS estimator was popularized by Angrist and Krueger (1992) and has been used successfully in several other empirical settings (e.g. (e.g., Bjorklund and Jantti, 1997; Currie and Yelowitz, 2000; Dee and Evans, 2003; Borjas, 2004).

## Appendix F

*Estimated Effects on Head Start Participation*

To get a sense of how our spending increases relate to changes in the Head Start participation margin, we used changes in national Head Start enrollment over time. However, given that Garces, Currie, and Thomas (2002) employ data on Head Start participation reported by PSID respondents, it is important to discuss the implied participation effects using these data. The data on Head Start participation used in Garces, Currie, and Thomas (2002) are imperfect in two important ways. First, the data are retrospective data collected in 1995 based on questions that asked adults about their Head Start participation at age 4 (several years before). Even though Garces, Currie, and Thomas (2002) present some evidence that any recall bias in these data may not be severe, we are reluctant to trust these data when there are other alternatives. The second major limitation of these data is that question about Head Start participation were only collected for the wave for individuals born after 1965 (i.e., who were 4 years old after 1969). As such, Head Start enrollment data were not recorded for the ramp-up period during which most of our variation is derived. This is further complicated by the fact that the largest increases in Head Start enrollment occurred between 1965 and 1970 in the Summer-only Programs, which were largely phased out 1970 onwards. As such, the large increases in Head Start participation (much of which was in the Summer only programs) between 1965 and 1970 are simply not recorded in the participation data recorded in the PSID. As a result, relating increases in Head Start spending to retrospectively reported Head Start enrollment in the PSID might drastically understate the effects of Head Start spending on enrollment in the program.

Having discussed all the limitations of using the reported Head Start enrollment in the PSID to infer the effects of spending on enrollment, it is helpful to show what estimates these data yield. About 19 percent of all poor children in our sample report attending Head Start (excluding summer programs) and about 33 percent of them had a Head Start center in their county at age 4. If we assume that all enrollees had a center in their county, this will imply that having any Head Start center in the county increases the likelihood of enrolling by 0.19/0.33= 57 percent. This is consistent with the national enrollment to low-income ratio after 1970 of between 55 and 65 percent. To explicitly model the relationship between increased spending on Head Start and the participation of low-income children in Head Start (using the self-reports from the PSID) we estimate conditional logit models. We predict Head Start participation using Head Start spending per poor four-year-old in the county while controlling for race/ethnicity and conditioning on the childhood county. We exclude controls for cohort trends because, by definition, such trends are zero before rollout. To allow for ease of interpretation, we report the average marginal treatment effects based on the conditional logit estimates. The marginal effects are presented in Table F1 for children from poor families.

The point estimates reveal that, for the poorest children, increasing Head Start spending by $1,000 per poor four-year-old would increase the likelihood of reporting enrollment in Head Start by 17.6 percentage points. This implies that for the average county that spends $4000 per poor four-year-old, Head Start participation is estimated to increase by 17.6*4=70.4 percentage points. Reassuringly, there is no effect of Head Start spending on enrolling in any non-Head Start Pre-Kindergarten program. This implied participation effect at rollout of 70 percentage points for poor children is very similar to our assumed participation effect of 75 percent for any Head Start

program, somewhat larger than our assumed participation effect of 50 percent for full-time Head Start, and of the same order of magnitude as both. Because these estimates are on a similar order of magnitude as those computed based on national data, we are confident that our preferred estimates of the participation margin from the national data are reasonably accurate.

**Table F1:**
*Conditional Logit Estimates of the Effects of Head Start Spending on Head Start Participation of Poor Children*

| | 1 | 2 |
|---|---|---|
| | Prob(Attend Head Start) | Prob(Attend Any Pre-K, but not Head Start) |
| | *Conditional Marginal Effects, evaluated at means* | |
| County Head Start Spending per Poor 4-year old$_{(age\ 4)}$ (in 000s) | 0.1765*** | 0.0175 |
| | (0.0202) | (0.0118) |
| Number of Children | 4,651 | 4,651 |
| Number of Childhood Families | 1,909 | 1,909 |
| Number of School Districts | 631 | 631 |
| Number of Childhood Counties | 448 | 448 |

Robust standard errors in parentheses (clustered at childhood state)
*** p<0.01, ** p<0.05, * p<0.1

Data: PSID geocode Data (1968-2013), matched with childhood school and neighborhood characteristics. Analysis sample includes PSID individuals born 1950-1976 who were followed through the 1995 survey IW. Child-specific pre-K attendance & Head Start program participation information collected retrospectively in 1995 survey IW. Poor children are those whose parents were in the bottom quartile of the income distribution (approximately 80% of whom were below the poverty line).

Models: Results are based on models that include school district fixed effects and controls for race/ethnicity.

## Appendix G

*Robustness Checks and Tests of Validity*

Because one of the parameters of interest is the marginal effect of the interaction between Head Start Spending and public K12 spending, it is important for us to establish that the variation we use in each of these is exogenous and will yield causal relationships. Here we present a series of empirical tests that support the validity of each source of variation.

*Head Start Spending Effects by Child Age*: *No confounding policies*. As a falsification/placebo test, we investigate the effects of Head Start spending increases by the child's age at which these increases occur. If the results are consistent with a causal interpretation of Head Start spending, then we would expect to find significant effects of that spending only for children who are age-eligible (age 4), and not for children who were already school-age at the time of the spending increase. Furthermore, even though our models control for a variety of other policies and we find no Head Start effects on non-poor children, one may still worry that the timing of Head Start rollout or the timing of SFRs coincided with other policies that also improved adult outcomes. One test of this would be to determine whether the effects of the spending increases are experienced only among those who were of the appropriate age. If counties or districts adopted other policies to improve outcomes for low-income children (that were not targeted to the exact same age range as that in question) one would observe improvements for other age ranges also. To test this for Head Start spending, we estimated the marginal effect of the level of Head Start spending that prevailed when the individual was different ages. To test whether Head Start spending at other ages predicts student outcomes, conditional on Head Start spending at age 4, we estimate the following regressions where all variables are as defined in [4] and [6].

[G1] $$Y_{icb} = \beta_w \cdot HS_{cb}^{age\,W} + \beta \cdot HS_{cb}^{age\,4} + \gamma \cdot C_{icb} + \theta_c + \tau_b + \varepsilon_{icb}.$$

We estimate models such as [G1] where we include our regressor of interest ($\beta_w$), the marginal effect of Head Start spending at age W on individual outcomes, conditional on the effect of Head Start spending at age 4. In principle, one should see that Head Start sending per poor four-year old has effects when the individual was four years old but not at other ages. This is exactly what we find across every one of the adult outcomes of poor children we analyze. In Figure 6 and Figure G1, we plot the marginal effect of Head Start spending by age conditional on spending at age 4. Note that the estimated effect for age 4 is not conditional on spending at other ages. However, the marginal effect of spending at age 4 is largely the same in models that include spending at other ages. The figures all show that increases in the Head Start spending level that prevailed when the individual was four years old are associated with significantly improved adult outcomes while the corresponding spending level at ineligible ages (1-3;5-10) are not.

Even though we instrument for K12 spending levels, it is important to establish that the identifying variation we use is valid. If the spending increases we exploit operate through improved K12 education, one should see improvement for those who were between the ages of 5 and 17 when there was a school finance reform, but no effect for individual from the same districts who were 18 or older at the time. Figure 5 shows that only those individuals who were of school-going age at the time of a reform-induced spending increase experience improved outcomes. These figures also reveal that outcomes in districts that saw increases in K12 spending were not on a positive or negative trajectory – indicating that the timing of the SFR was exogenous to the underlying trends in outcomes in affected districts. To show this more formally, we estimate models that instrument for K12 spending in an individual's childhood districts when they were

between the ages of 20 and 24. Results are in appendix Table G1. If the effects are real, we should see effects for reform-induced spending increases when an individual was between the ages of 5 and 17 but not for increases that occurred when an individual was between the ages of 20 and 24. As in Jackson, Johnson and Persico (2016), K12 spending levels between ages 20 to 24 have no effect on outcomes.

*Instrumenting for Head Start Spending*

Even though we present several empirical tests that suggest that the within county variation in Head Start identified in [1] is valid, one may worry that some of the variation in Head Start spending is determined by county-level choices that may be endogenous to student outcomes. For example, counties that decided to implement generous Head Start funding levels may also have implemented other policies that affected student outcomes or may differ in unobserved ways from those that decided on less generous funding levels. Though we find no evidence that this was the case, to address this possibility we proposed an instrumental variables strategy that removes the variation in the level of Head Start spending that is driven by county level decisions and relies exclusively on variation in the timing of Head Start rollout.

Figure 4 shows that most of the within-county variation in Head Start spending over time occurs before and after rollout. As such, we instrument for the level of Head Start spending per poor 4-year old in an individual's county when they were 4 years old with an indicator for whether Head Start was available in the individual's childhood county when they were 4 years old, and the number of years a Head Start center had existed in an individual's childhood county at age 4. This instrumental variables strategy compares the adult outcome of individuals who were 4 years old before Head Start was available in their county (unexposed cohorts with zero spending), to the outcomes of adults from the same childhood county who were 4 years old soon after rollout (early exposed cohort with moderate levels of Head Start spending, on average) and also those from the same childhood county who were 4 years old long after rollout (late exposed cohorts with high levels of Head Start spending, on average). To account for underlying cohort level differences on can use the difference across the same birth cohorts in non-rollout counties as a basis for comparison. The identifying assumption is that the timing of Head Start rollout in a county is unrelated to other changes at the county level that exert an independent effect on adult outcomes. Under this identifying assumption, if, *among individuals from the same childhood county*, the late exposed cohorts have better outcomes than the early exposed cohorts, and the early exposed cohorts have better outcomes than the unexposed cohorts (relative to the differences across the same birth cohorts in counties that did not rollout Head Start across the same time period), it would imply a real casual effect of Head Start spending on adult outcomes. Figure 4 already presents some visual evidence of this.

The Instrumental Variables model that exploits the variation described above is implemented by estimating [G2] and [G3] by Two Stage Least Squares (2SLS) where [G2] is a first stage regression based on the Head Start rollout instruments.

[G2] $\widehat{HS}_{icb}^{age\,4} = \rho_1 \cdot HSExp_{icb} + \rho_2 \cdot Years(HSExp_{icb}) + \rho_3 \cdot C_{icb} + \theta_{c1} + \theta_{b1} + \varphi_{icb}.$

[G3] $Y_{icb} = \beta \cdot \widehat{HS}_{icb}^{age\,4} + \gamma \cdot C_{icb} + \theta_c + \theta_b + \varepsilon_{icb}.$

All common variables are defined as in [3]. $\widehat{HS}_{cb}^{age\,4}$ is the first stage prediction of Head Start spending per 4-year old, $HSExp$ is an indicator for whether individual *i* from birth cohort *b* was age four (or younger) when Head Start was first established in their childhood county *c*, and $Years(HSExp_{icb})$ is the number of years that Head Start had been established in childhood county *c* for individual *i*, at the time birth cohort b was 4 years old.

To also test our interaction model, we implement models that instrument for both K12 spending and the interaction, and also for Head Start spending. Specifically, we estimate [G4] below, where $\widehat{HS}_{cb}^{age\,4}$ is estimated in a first stage using the number of years that Head Start existed in an individual's county as our exogenous predictor of Head Start spending per four-year-old at age 4, and $\widehat{INT}_{idb}$ is estimated in a first stage based on the interactions between the Head Start rollout instruments and the SFR-reforms instruments.

[G4] $$Y_{icb} = \beta_{HS} \cdot \widehat{HS}_{cb}^{age\,4} + \beta_{k12} \cdot \ln\left(\widehat{PPE}_{5-17}\right)_{idb} + \beta_{int} \cdot (\widehat{INT}_{idb}) + \gamma \cdot C_{icb} + \theta_d + \theta_b + \varepsilon_{idb}.$$

In [G4], all of the exogenous shifters in K12 spending ($(SFRExp_{idb} \times dose_d)$ and $(SFRExp_{idb})$) are interacted with the exogenous shifters of Head Start spending ($(HSExp_{icb})$ and $Years(HSExp_{icb})$), and are used to create the first stage predictions of K12 spending ($\ln\left(\widehat{PPE}_{5-17}\right)_{idb}$), Head Start spending ($\widehat{HS}_{cb}^{age\,4}$) and the interaction between the two ($\widehat{INT}_{idb}$). While this model yields noisier estimates, if the results between model [G4] are similar to those in [6] it will be compelling evidence that our estimated effects are real.

Results from this strategy that instruments for K12 spending, Head Start spending and the interactions between the two is presented in Table G2. While this model yields less precise point estimates than those in Table 2, the pattern of results is very similar – this suggest that changes in Head Start spending are unrelated to other changes. Because these 2SLS models that use the timing of Head Start rollout rely on the exogeneity of the timing of Head Start rollout, we also present evidence that this alternate variation is likely valid. To show this, we present event study figures for the effect of Head Start rollout for counties with high spending level versus other counties in Figure 4. As discussed previously, there is no evidence of any pre-existing time trends, suggesting that the timing of rollout was exogenous.

*No selection or endogenous mobility:* Another concern one may have with the estimates is that due to selective migration or neighborhood change, the characteristics of the individuals exposed to different levels of K12 spending or Head Start spending are not the same. We address this possible concern in two ways. First, we demonstrate that the spending changes we exploit are unrelated to observed family and neighborhood characteristics. Specifically, we regress year of educational attainment and the adult wage on several observable characteristics and then take the fitted values from those regressions as our predicted outcomes. To obtain these predicted outcomes, we estimated models that predict educational attainment and adult earnings using parental income, race, mother's and father's education and occupational prestige index, mother's marital status at birth, birth weight, childhood county-level average per-capita expenditures on Title I, AFDC, Medicaid, food stamps, and UI, respectively, during childhood years. The predicted outcomes from these models are intended to capture an effect-size weighted index of childhood family/community SES factors. We then regress our predicted outcomes on the spending changes (excluding all of these same observable characteristics). If the spending changes are unrelated to those observable characteristics that predict the adult outcome, the estimated coefficients will be zero. Indeed, this is what we find (See Table G3).

Even though our spending changes are unrelated to observed characteristics, one may worry about selection on unobserved characteristics. To rule out the possibility that our results are driven by differences across treated and untreated families, we rely on variation within families and compare the outcomes of siblings who were different ages at Head Start rollout or at the time of a court-ordered SFR, but were raised in the same household. This approach accounts for all observed and unobserved shared family characteristics that predict outcomes. We achieve this by

augmenting [6] to include sibling fixed effects. As one can see in Table G4, the estimated coefficients for low-income children are very similar to those without sibling fixed effects. This suggests that family selection cannot explain the main pattern of results. This sibling tests also addresses any potential lingering concerns regarding endogenous mobility driving the results because individuals in the same family have the same residential address. As an additional check on endogenous mobility, we re-estimated all models limiting the analysis sample to those who lived at their (earliest) childhood residence before the enactment of Head Start programs in their respective county. The results are presented in Appendix Table G5. We find nearly identical results as those in the full sample. This indicates that endogenous residential mobility is not a major source of bias in this analysis.

*Testing for Sufficient Variation to Identify the Interaction Term*

Identification of our parameter of interest is based on the interaction between two policy instruments. Credible identification of our parameter requires that there be exogenous variation in both Head Start spending and K12 spending conditional on the other. This issue is discussed in Buckles, Morrill, Hagerman, Wozniak and Malamud (2013). Intuitively, if the same areas that receive increased K12 spending due to reforms are also those that experienced the largest increases in Head Start spending, then there may be no credible exogenous variation in K12 spending conditional on Head Start spending and *vice versa*. With a very high correlation between the two policy instruments, our model would be underidentifed.

We assess whether this is a problem in two different ways. First, we compute the correlation between Head Start spending per poor 4-year old (at age 4) and instrumented ln(K12 spending) at the childhood county-birth cohort level. If our policy-induced variation in Head Start spending and K12 spending were based on the same sample of counties, there would be a large positive correlation. In fact, the raw correlation between Head Start spending per poor 4-year old (at age 4) and instrumented ln(K12 spending) is only 0.0844. To test this more formally, we ran our 2SLS model predicting Head Start spending at age four as a function of the SFR-induced changes in K12 spending with all the controls from our main specification. The results are presented in Table G6. In such models, the coefficient is an economically insignificant -0.33 and the *p*-value is larger than 0.1. Taken at face value, the point estimate indicates that an exogenous 10 percent increase in K12 spending is associated with a mere additional $30 per poor four-year-old spent on Head Start. Similarly, we regressed the reform-induced change in K12 spending (the fitted values from the first-stage regression predicting K12 spending) on Head Start spending at age 4. In such models, the coefficient is less than 0.0001 with a *p*-value greater than 0.1. In sum, the two sources of exogenous variation are largely unrelated to one another, such that the interaction between the two is identified.

As a further check that there is sufficient variation to uniquely identify each of our endogenous regressors, we follow Angrist and Pischke (2009). To test for sufficient unique variation in our main models that rely on difference-in-difference variation in Head Start spending and instrument for both K12 spending and the interaction between K12 spending and Head Start spending, we report a series of F-statistics (see Table G7). Looking at predicting K12 spending, the first stage F-statistic for the log of K12 spending (based on predicted district-level dosage times years of SFR exposure in the state) is 51.85 and 40.86 in models without and with Head Start variables, respectively. As such, there is a strong first stage for K12 spending whether Head Start spending is included in the model or not. Also, as a direct test of the strength of the first stage for the interaction in our main models, the first stage F-statistic for Head Start spending times SFR

dosage times SFR exposure is 23.05, conditional on Head Start spending and SFR dosage times SFR exposure. That is, the F-statistic on the interaction between the two policy instruments in predicting the interaction between the two spending types is large (conditional on the effect of the individual policy instruments themselves). In sum, all the tests presented indicate that all of the endogenous regressors in our main model can be identified.

Given that one of our robustness checks instruments for Head Start spending using the timing of the rollout of Head Start in an individual's childhood county, it is helpful to show that this double instrumental variables model can also be identified. The first stage F-statistic for Head Start spending (based only on the timing of rollout as the excluded instrument) is 23.25 in models with no K12 variables included. In models with the SFR instruments for K12 spending, the first stage F-statistic of the Head Start instruments is very similar at 25.72. Finally, in predicting the interaction between both spending types, the first stage F-statistic for Head Start *exposure* times SFR dosage times SFR exposure is 18.45, conditional on Head Start *exposure* and SFR dosage times SFR exposure. In sum, all the tests indicate that we have sufficient independent exogenous variation to credibly identify the effects of Head Start spending, the effect of K2 spending, and the effects of the interaction between the two.

**Appendix Table G1:**

*2SLS/IV Estimates of Court-Ordered School Finance Reform Induced Effects of Per-Pupil Spending on Long-Run Outcomes: Placebo Tests for Non-school Ages* (All children. All adult outcomes are measured between ages 20-45)

| | Years of Education | Prob(High School Grad) | Ln(Wage) | Ln(Family Income) | Prob(poverty) | Prob(incarcerated) |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | | 5 |
| $Ln(PPE_d)_{(age\ 5\text{-}17)}$ | 3.9461+ | 0.4387+ | 1.3782+ | 2.4754+ | -0.6717+ | -0.0253 |
| | (2.501) | (0.3326) | (0.8598) | (1.5470) | (0.5092) | (0.2543) |
| $Ln(PPE_d)_{(age\ 20\text{-}24)}$ | -1.3684 | 0.2331 | -0.5722 | -1.5110 | 0.4552 | -0.1013 |
| | (1.5089) | (0.2507) | (0.5190) | (1.0420) | (0.3413) | (0.1412) |
| Number of person-year observations | | | 49,282 | 77,945 | 78,213 | |
| Number of Individuals | 5,385 | 5,378 | 4,981 | 5,621 | 5,623 | 4,317 |
| Number of Districts | 762 | 761 | 760 | 782 | 782 | 528 |

Robust standard errors in parentheses (clustered at childhood stat level) *** p<0.01, ** p<0.05, * p<0.10

Data: PSID geocode Data (1968-2011), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1955-1985, followed into adulthood through 2011. Sampling weights are used so that the results are nationally representative.

Models: The key treatment variable, $Ln(PPE_d)_{(age\ 5\text{-}17)}$, is the natural log of average school-age per-pupil spending. All models include school district fixed effects, birth cohort fixed effects, and the additional controls listed below. The excluded instruments from the second stage are (the number of years of exposure to a court-ordered SFR) and (the number of years of exposure to a court-ordered SFR) × (the quartile of the district in the distribution of $Spend_d$) and (the number of years of between the ages of 20 and 24 that occur after a court-ordered SFR) and (the number of years of between the ages of 20 and 24 that occur after a court-ordered SFR) × (the quartile of the district in the distribution of $dose_d$).

Additional controls: childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). Also race × census division × birth cohort fixed effects; controls at the county-level for the timing of school desegregation by race, hospital desegregation × race, rollout of community health centers, county expenditures on Head Start (at age 4), food stamps, medicaid, AFDC, UI, Title-I (average during childhood years), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends.

**Table G2:**

*2SLS-2SLS Estimates of Early and K12 education Spending on Adult Outcomes*

| | Children from Poor Households | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| | Prob(High School Grad) | Years of Education | Ln(Wage), ages 20-50 | Ln(Family Income), ages 20-50 | Poverty, ages 20-50 | Prob(Ever Incarcerated) |
| Instrumented: Head Start Spending$_{(age\ 4)}$/1000 | 0.09010*** | 0.2768** | 0.03779* | 0.05518** | -0.02969*** | -0.01936* |
| | (0.03177) | (0.1236) | (0.01966) | (0.02657) | (0.01146) | (0.01117) |
| Instrumented: Ln(K12 Per-pupil Spending)$_{(age\ 5\text{-}17)}$ | 1.5488*** | 3.4924* | 0.8023* | 1.5388** | -0.4070+ | -1.5206** |
| | (0.4422) | (2.0229) | (0.4570) | (0.7830) | (0.3084) | (0.7175) |
| Instrumented: Interaction | 0.3207*** | 0.4744* | 0.1574** | 0.2732*** | -0.1534*** | -0.1411* |
| | (0.1087) | (0.2587) | (0.06900) | (0.1010) | (0.05388) | (0.08434) |
| Number of Person-year Observations | | | 49,282 | 54,064 | 54,309 | |
| Number of Childhood Families | 2,024 | 2,023 | 2,016 | 2,098 | 2,098 | 1,693 |
| Number of Children | 5,385 | 5,378 | 5,378 | 5,378 | 5,378 | 4,317 |

Robust standard errors in parentheses (clustered at childhood state level)
*** p<0.01, ** p<0.05, * p<0.1

Data: PSID geocode Data (1968-2013), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 whose parents were in the bottom quartile of the income distribution, and who have been followed into adulthood.

Models: Instrumented Head Start spending per poor 4-year old in the county is centered around $4,230 (and measured in 000s) and instrumented ln(school district per-pupil spending during ages 5-17) is centered around 1.6, to facilitate interpretation of the main effects evaluated at roughly the respective means; the average SFR-induced increase in school-age spending is about 10%. Results are based on 2SLS-IV models that include: parent's relative rank in income distribution, school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, rollout of "War on Poverty" & related safety-net programs (community health centers, food stamps, medicaid, AFDC, UI, Title-I (average during childhood yrs)), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends; and controls for childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender), and age (cubic). The first-stage model include as predictors the school-age years of exposure to school finance reform interacted with the quartile of the respective school district's predicted reform-induced change in school spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories. The instruments for Head Start spending per poor 4-year old include: a dummy indicator of whether a county Head Start center had ever been established by the time the individual was age 4 and the number of years a Head Start center had existed in childhood county at age 4; and these instruments are interacted with school-age years of exposure to school finance reform*top quartile of the respective school district's predicted reform-induced change in school spending based on the timing and type of court-ordered reform. There exists a significant first-stage.

**Table G3.**

*Examining Exogeneity of Head Start and K-12 Spending (Low Income Children)*

| | Predicted Years of Education, based on Childhood Family & County SES | | Predicted Ln(Wages) at age 30, based on Childhood Family & County SES | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| | School District FE & Race*Birth Yr FE | Partial Set of Controls | School District FE & Race*Birth Yr FE | Partial Set of Controls |
| Head Start Spending$_{(age\ 4)}$/1000 | -0.0034853 | -0.0000969 | 0.0005169 | 0.0007319 |
| | (0.0034095) | (0.0039868) | (0.0006791) | (0.0006827) |
| Ln(K12 Per-pupil Spending)$_{(age\ 5\text{-}17)}$ | -0.383190 | -0.3197846 | 0.0313539 | -0.0175643 |
| | (0.667197) | (0.6115767) | (0.1124003) | (0.1132423) |

*** p<0.01, ** p<0.05, * p<0.10; Robust standard errors in parentheses (clustered at childhood state level)

Data: PSID geocode Data (1968-2013), matched with childhood school and neighborhood characteristics. Analysis sample includes PSID individuals born 1950-1976, followed into adulthood through 2013. We estimated models that predict educational attainment & adult earnings using only childhood family/community SES characteristics (including parental income, race, mother's and father's education and occupational prestige index, mother's marital status at birth, birth weight, childhood county-level average per-capita expenditures on Title I, AFDC, Medicaid, food stamps, & UI, respectively, during childhood years)—this is intended to capture an effect-size weighted index of childhood family/community SES factors. We then examined whether individuals' predicted educational attainment, likelihood of high school graduation, and wages at age 30 based only on childhood family/county characteristics (i.e., the effect-size weighted index of childhood family/community SES factors) is related to county Head Start spending per poor 4-year old, holding constant school district fixed effects and birth year fixed effects. Head Start spending per poor 4-year old is in thousands of dollars (real 2000 dollars), so that a one-unit change represents a $1,000 change in spending. Partial controls include neighborhood controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, rollout of community health centers, timing of state-funded Kindergarten intro and timing of tax limit policies. The partial controls also include race-specific year of birth fixed effects, race*census division-specific birth year trends and controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size) each interacted with linear cohort trends. The models with no additional controls show no correlation between the policy instruments and those childhood family and neighborhood characteristics that predict the outcomes. Models with partial controls show no correlation between the policy instruments and those family characteristics that predict the outcomes.

**Table G4:**

*Within family Model: 2SLS-Difference-in-Difference Estimates of Early and K12 education Spending on Adult Outcomes*

| | Children from Poor Households | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| | Prob(High School Grad) | Years of Education | Ln(Wage), ages 20-50 | Ln(Family Income), ages 20-50 | Poverty, ages 20-50 | Prob(Ever Incarcerated) |
| Head Start Spending$_{(age\ 4)}$/1000 | 0.02044* | 0.1153*** | 0.008247+ | 0.02255** | -0.002645 | -0.01890*** |
| | (0.01096) | (0.02779) | (0.005471) | (0.009970) | (0.003889) | (0.006655) |
| Ln(K12 Per-pupil Spending)$_{(age\ 5\text{-}17)}$ | 0.9831** | 4.4541*** | 0.7861+ | 2.3433** | -0.4527* | -1.3356* |
| | (0.4508) | (1.7092) | (0.5467) | (0.9161) | (0.2667) | (0.7545) |
| Interaction | 0.09013+ | 0.5500** | 0.01816 | 0.09604 | -0.05668+ | 0.03364 |
| | (0.06069) | (0.2457) | (0.05940) | (0.09992) | (0.04317) | (0.07735) |
| Number of Person-year Observations | | | 49,282 | 54,064 | 54,309 | |
| Number of Childhood Families | 2,024 | 2,023 | 2,016 | 2,098 | 2,098 | 1,693 |
| Number of Children | 5,385 | 5,378 | 5,378 | 5,378 | 5,378 | 4,317 |

Robust standard errors in parentheses (clustered at childhood state level)

*** p<0.01, ** p<0.05, * p<0.1

Data: PSID geocode Data (1968-2013), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 whose parents were in the bottom quartile of the income distribution, and who have been followed into adulthood.

Models: Head Start spending per poor 4-year old in the county is centered around $4,230 (and measured in 000s) and instrumented ln(school district per-pupil spending during ages 5-17) is centered around 1.6, to facilitate interpretation of the main effects evaluated at roughly the respective means; the average SFR-induced increase in school-age spending is about 10%. Results are based on 2SLS-Difference-in-Difference models that include: parent's relative rank in income distribution, school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, rollout of "War on Poverty" & related safety-net programs (community health centers, food stamps, medicaid, AFDC, UI, Title-I (average during childhood yrs)), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends; and controls for childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). The first-stage model include as predictors the school-age years of exposure to school finance reform interacted with the quartile of the respective school district's predicted reform-induced change in school spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories. There exists a significant first-stage.

**Table G5:**

*Early Address Sample:*

*2SLS-Difference-in-Difference Estimates of Early and K12 education Spending on Adult Outcomes*

| | Children from Poor Households | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| | Prob(High School Grad) | Years of Education | Ln(Wage), ages 20-50 | Ln(Family Income), ages 20-50 | Poverty, ages 20-50 | Prob(Ever Incarcerated) |
| Head Start Spending$_{(age\ 4)}$/1000 | 0.02142*** | 0.09016*** | 0.01745*** | 0.03455*** | -0.009435*** | -0.007252** |
| | (0.005316) | (0.02651) | (0.004118) | (0.005861) | (0.002620) | (0.0033937) |
| Ln(K12 Per-pupil Spending)$_{(age\ 5\text{-}17)}$ | 0.7990* | 4.0147** | 1.4607+ | 0.1984 | -0.05884 | -1.471573** |
| | (0.4791) | (2.0424) | (0.9832) | (1.5506) | (0.5108) | (0.581283) |
| Interaction | 0.1344* | 0.5743+ | 0.1784*** | 0.1054+ | -0.03840+ | -0.0727697* |
| | (0.07012) | (0.3795) | (0.04012) | (0.07024) | (0.02859) | (0.0386209) |
| Number of Person-year Observations | -- | -- | 47,804 | 74,654 | 74,909 | -- |
| Number of Children | 5,006 | 4,999 | 4,685 | 5,263 | 5,265 | 4,119 |

Robust standard errors in parentheses (clustered at childhood state level)

*** p<0.01, ** p<0.05, * p<0.1

Data: PSID geocode Data (1968-2013), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 whose parents were in the bottom quartile of the income distribution, and who have been followed into adulthood.

Models: Head Start spending per poor 4-year old in the county is centered around $4,230 (and measured in 000s) and instrumented ln(school district per-pupil spending during ages 5-17) is centered around 1.6, to facilitate interpretation of the main effects evaluated at roughly the respective means; the average SFR-induced increase in school-age spending is about 10%. Results are based on 2SLS-Difference-in-Difference models that include: parent's relative rank in income distribution, school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, rollout of "War on Poverty" & related safety-net programs (community health centers, food stamps, medicaid, AFDC, UI, Title-I (average during childhood years)), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends; and controls for childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender). The first-stage model include as predictors the school-age years of exposure to school finance reform interacted with the quartile of the respective school district's predicted reform-induced change in school spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories. There exists a significant first-stage.

**Table G6:**

*The Relationship between School Finance Reform-Induced Changes in Per-Pupil K12 Spending and Head Start Spending*

(**Children from Poor Households**)

| | **Dependent variable:** | |
|---|---|---|
| | (1) | (2) |
| | County Head Start Spending per Poor 4-year old$_{(age\ 4)}$ (in 000s) | (SFR) Instrumented Ln(School District Per-pupil Spending)$_{(age\ 5\text{-}17)}$ |
| County Head Start Spending per Poor 4-year old$_{(age\ 4)}$ (in 000s) | -- | 0.00000494 |
| | | (0.0004672) |
| (SFR) Instrumented Ln(School District Per-pupil Spending)$_{(age\ 5\text{-}17)}$ | -0.3302 | -- |
| | (1.8491) | |
| Number of Children | 5,385 | 5,385 |

Robust standard errors in parentheses (clustered at childhood state level)

*** p<0.01, ** p<0.05, * p<0.1

Data: PSID geocode Data (1968-2013), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 whose parents were in the bottom quartile of the income distribution, and who have been followed into adulthood.

At the childhood county-birth cohort level, the correlation between Head Start spending per poor 4-year old (at age 4) and instrumented ln(K12 spending) is 0.0844; and controlling for birth year, there is no significant relationship.

Models: Head Start spending per poor 4-year old in the county is centered around $4,230 (and measured in 000s) and instrumented ln(school district per-pupil spending during ages 5-17) is centered around 1.6, to facilitate interpretation of the main effects evaluated at roughly the respective means; the average SFR-induced increase in school-age spending is about 10%. Results are based on 2SLS-Difference-in-Difference models that include: parent's relative rank in income distribution, school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, rollout of "War on Poverty" & related safety-net programs (community health centers, food stamps, medicaid, AFDC, UI, Title-I (average during childhood yrs)), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends; and controls for childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender), and age (cubic). The first-stage model include as predictors the school-age years of exposure to school finance reform interacted with the quartile of the respective school district's predicted reform-induced change in school spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories. There exists a significant first-stage.

**Table G7:**

*F-Statistics on Excluded Instrument in Different Models (Poor children only)*

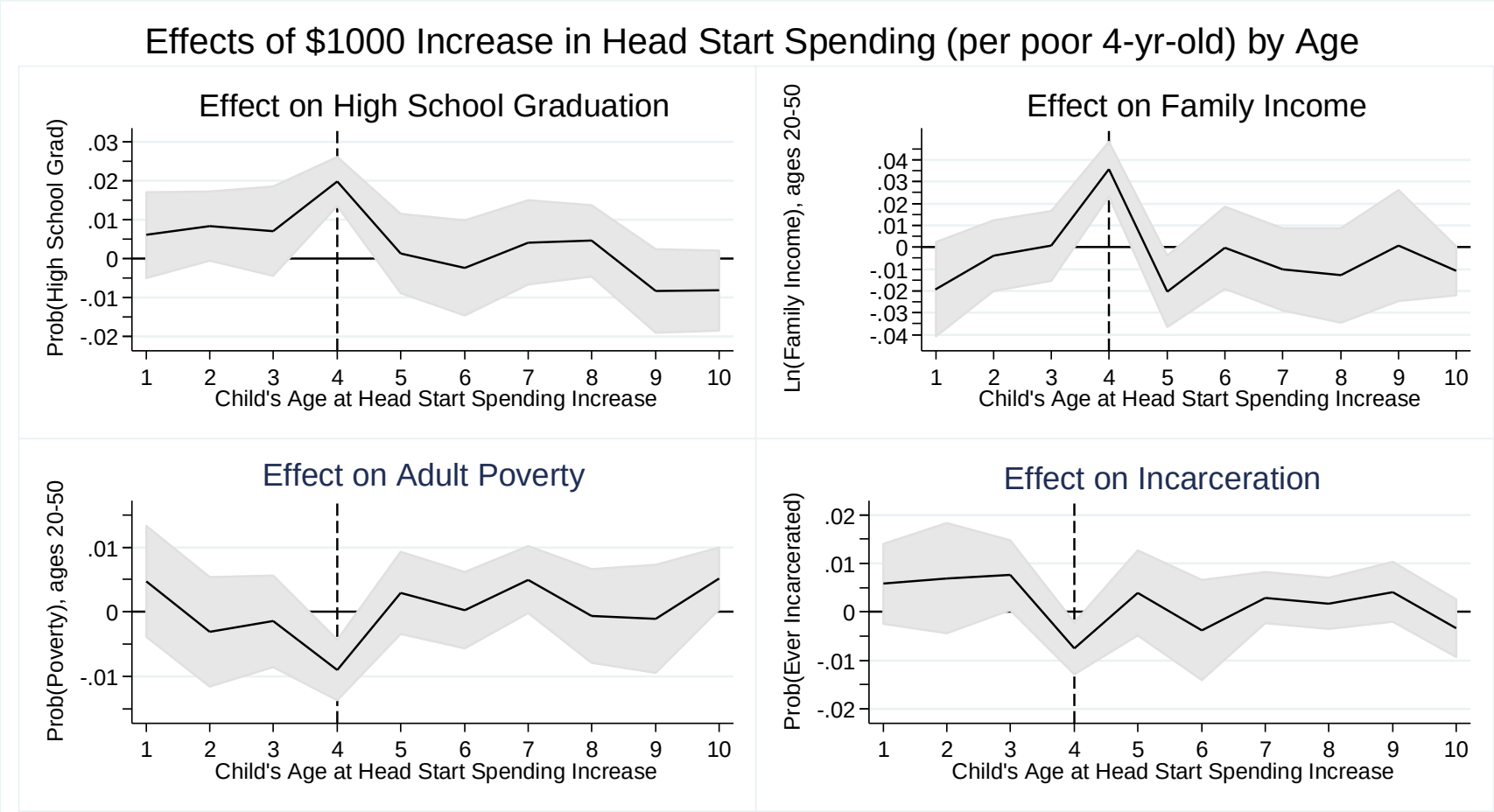
| | (3) | (4) | (5) | (6) | (1) | (2) |
|---|---|---|---|---|---|---|
| Dependent Variable | Ln(School District Per-pupil Spending)$_{(age\ 5\text{-}17)}$ | | Head Start Spending*Ln(K12 Spending) | | County Head Start Spending per Poor 4-year old$_{(age\ 4)}$ | |
| Model | without Head Start Spending | with Head Start Spending | With Head Start Spending and SFR dosage*SFR exposure | With Head Start Exposure and SFR dosage*SFR exposure | without SFR dosage*SFR exposure | with SFR dosage*SFR exposure |
| Excluded Instruments | SFR dosage*SFR exposure | SFR dosage*SFR exposure | Head Start Spending*SFR dosage*SFR exposure | Head Start Exposure*SFR dosage*SFR exposure | Head Start Exposure | Head Start Exposure |
| F-Statistic on excluded instruments | 51.85 | 40.86 | 23.05 | 18.45 | 23.25 | 25.72 |

Robust standard errors clustered at childhood state level.

Data: PSID geocode Data (1968-2013), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 whose parents were in the bottom quartile of the income distribution, and who have been followed into adulthood.

Models include: parent's relative rank in income distribution, school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, rollout of "War on Poverty" & related safety-net programs (community health centers, food stamps, medicaid, AFDC, UI, Title-I (average during childhood yrs)), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends; and controls for childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender), and age (cubic). The first-stage model of K12 spending include as predictors the school-age years of exposure to school finance reform interacted with the quartile of the respective school district's predicted reform-induced change in school spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories. The instruments for Head Start spending per poor 4-year old include: a dummy indicator of whether a county Head Start center had ever been established by the time the individual was age 4 and the number of years a Head Start center had existed in childhood county at age 4; and in column (2) these instruments are interacted with school-age years of exposure to school finance reform*top quartile of the respective school district's predicted reform-induced change in school spending based on the timing and type of court-ordered reform.

**Figure G1:**
*Effect of Head Start Spending by Age (other outcomes)*




These figures present the marginal effects of Head Start spending in an individual's childhood county at different ages, conditional on the level of Head Start spending in the childhood county at age 4 (when such spending should have an effect). The sample is poor children only. Models include the set full set of controls as in Tables 2 and 3. The coefficients on the non-eligible years 1 through 3 and 5 through 10, are all conditional on spending at age 4. The coefficient for spending at age 4 is based on model with no other ages included.

**Appendix Table G6.** *Placebo Tests: Effects of Head Start Spending by Child Age on Educational Attainment, Low-Income Children*

| | Completed Years of Education | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Age 4: County Head Start Spending per Poor 4-year old$_{(age\ 4)}$ (in 000s) | 0.09679*** | 0.09485*** | 0.09775*** | 0.09122*** | 0.09078*** | 0.09720*** | 0.09575*** | 0.09749*** | 0.09996*** | 0.09876*** |
| | (0.01912) | (0.01921) | (0.02547) | (0.02060) | (0.01884) | (0.02227) | (0.01854) | (0.01860) | (0.01879) | (0.01931) |
| Age1: County Head Start Spending per Poor 4-year old$_{(age\ 1)}$ (in 000s) | | 0.02183 | | | | | | | | |
| | | (0.02631) | | | | | | | | |
| Age2: County Head Start Spending per Poor 4-year old$_{(age\ 2)}$ (in 000s) | | | -0.002210 | | | | | | | |
| | | | (0.03653) | | | | | | | |
| Age3 County Head Start Spending per Poor 4-year old$_{(age\ 3)}$ (in 000s) | | | | 0.02781 | | | | | | |
| | | | | (0.02008) | | | | | | |
| Age5: County Head Start Spending per Poor 4-year old$_{(age\ 5)}$ (in 000s) | | | | | 0.02093 | | | | | |
| | | | | | (0.02581) | | | | | |
| Age6: County Head Start Spending per Poor 4-year old$_{(age\ 6)}$ (in 000s) | | | | | | -0.002726 | | | | |
| | | | | | | (0.02477) | | | | |
| Age7: County Head Start Spending per Poor 4-year old$_{(age\ 7)}$ (in 000s) | | | | | | | 0.003822 | | | |
| | | | | | | | (0.02156) | | | |
| Age8: County Head Start Spending per Poor 4-year old$_{(age\ 8)}$ (in 000s) | | | | | | | | -0.003993 | | |
| | | | | | | | | (0.01848) | | |
| Age9: County Head Start Spending per Poor 4-year old$_{(age\ 9)}$ (in 000s) | | | | | | | | | -0.03386 | |
| | | | | | | | | | (0.02795) | |
| Age10: County Head Start Spending per Poor 4-year old$_{(age\ 10)}$ (in 000s) | | | | | | | | | | -0.006153 |
| | | | | | | | | | | (0.01650) |
| Number of Children | 5,378 | 5,378 | 5,378 | 5,378 | 5,378 | 5,378 | 5,378 | 5,378 | 5,378 | 5,378 |
| Number of School Districts | 761 | 761 | 761 | 761 | 761 | 761 | 761 | 761 | 761 | 761 |
| Number of Childhood Counties | 577 | 577 | 577 | 577 | 577 | 577 | 577 | 577 | 577 | 577 |

Robust standard errors in parentheses (clustered at childhood county level)

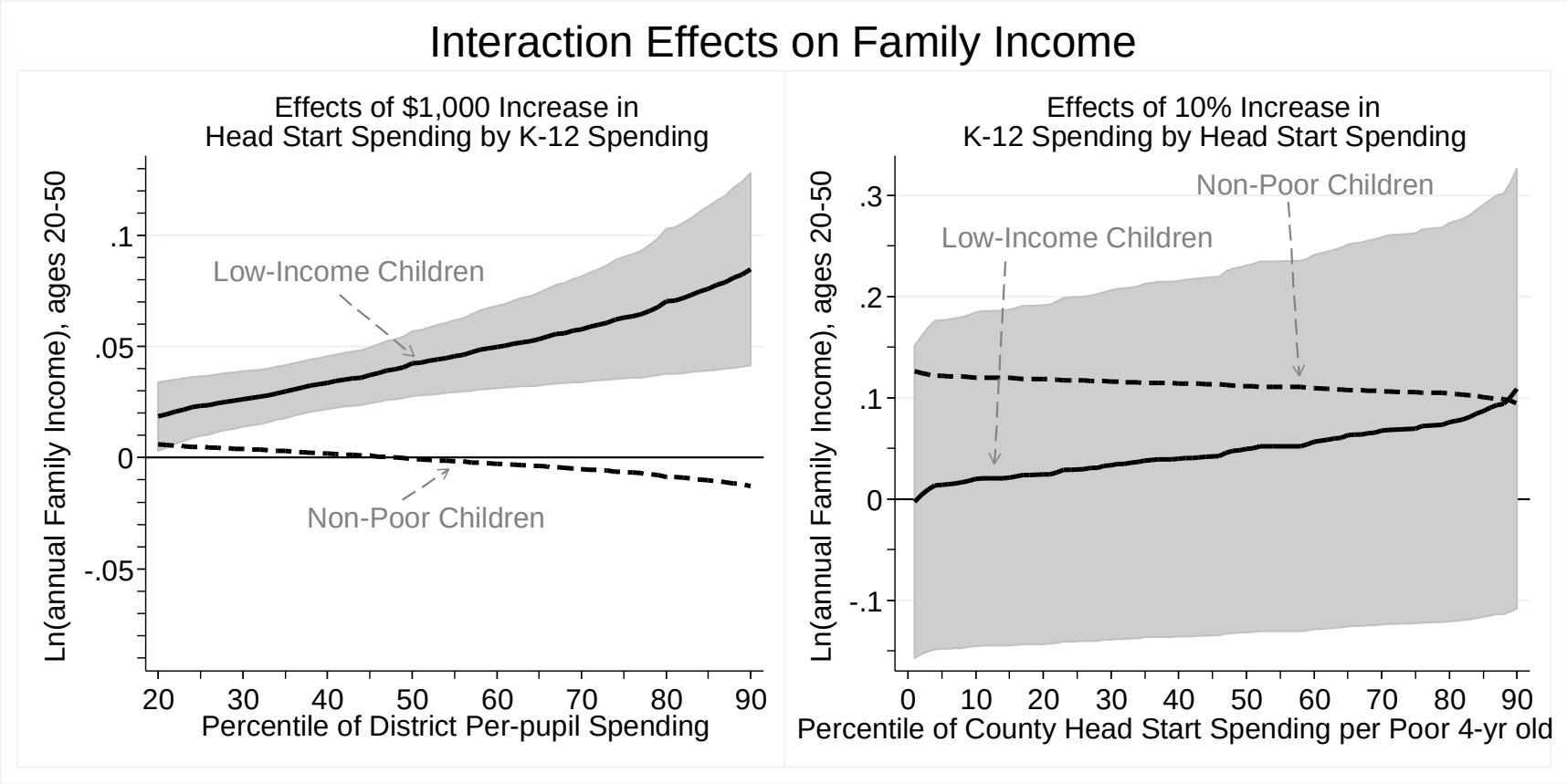
*** p<0.01, ** p<0.05, * p<0.1

Data: PSID geocode Data (1968-2013), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 whose parents were in the bottom quartile of the income distribution, and who have been followed into adulthood.

Models: Head Start spending per poor 4-year old in the county is measured in 000s. These results are also presented in Figures 7a-7f across all outcomes. Results are based on Difference-in-Difference models that include same full set of controls (as in Tables 1-2): parent's relative rank in income distribution, school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, rollout of "War on Poverty" & related safety-net programs (community health centers, food stamps, medicaid, AFDC, UI, Title-I (average during childhood years)), timing of state-funded Kindergarten introduction and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends; and controls for childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender).

**Appendix H**

**Figure H1:**
*Additional dynamic complementarity plots for family income*




Note: The reported marginal Effects based upon 2SLS-Difference-in-Difference model results presented in columns 4 from Tables 2 and 3. The reported marginal effects and the standard errors were computed using the delta method.

## Appendix I

*Testing for Improvement in Parent Quality due to Head Start*

**Table I1:**

*Test for any Spillover Effects of Head Start Spending on Older Sibling not Exposed at age 4*

| | Children from Low Income Families | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| | Prob(High School Grad) | Years of Education | Ln(Wage), ages 20-50 | Ln(annual Family Income), ages 20-50 | Annual Incidence of Poverty, ages 20-50 | Prob(Ever Incarcerated) |
| Younger Sibling's County Head Start Spending per Poor 4-year old$_{(age\ 4)}$ (in 000s) | -0.0215 | 0.0026 | 0.0129 | 0.0037 | 0.0063 | 0.00011 |
| | (0.0180) | (0.0685) | (0.0313) | (0.0639) | (0.0265) | (0.0045) |

Robust standard errors in parentheses (clustered at childhood state level)

*** p<0.01, ** p<0.05, * p<0.1

Data: PSID geocode Data (1968-2013), matched with childhood school and neighborhood characteristics. Analysis sample includes only older siblings not exposed to Head Start (i.e., who turned age 4 before the program's rollout), but whose younger sibling(s) had a Head Start center in the county when they were age 4, and whose parents were in the bottom quartile of the income distribution, and who have been followed into adulthood.

Models: Head Start spending per poor 4-year old in the county (measured in 000s). Results are based on models that include same set of controls as Tables 1-2: parent's relative rank in income distribution, school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends; controls at the county-level for the timing of school desegregation*race, hospital desegregation*race, rollout of "War on Poverty" & related safety-net programs (community health centers, food stamps (average during age 0-4), Medicaid (average during age 0-4), AFDC, UI, Title-I (average during childhood years)), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, percent voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends; and controls for childhood family characteristics (parental income/education/occupation, mother's marital status at birth, birth weight, gender), and age (cubic).