2019

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

Rucker C Johnson, *University of California, Berkeley*
C. Kirabo Jackson, *Northwestern University*

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

Rucker C. Johnson*
University of California, Berkeley
and NBER

C. Kirabo Jackson
Northwestern University
and NBER

December 2018

We compare the adult outcomes of cohorts who were differentially exposed to policy-induced changes in Head Start and K12 spending, depending on place and year of birth. IV and sibling-difference estimates indicate that, for poor children, these policies both increased educational attainment and earnings, and reduced poverty and incarceration. The benefits of Head Start were larger when followed by access to better-funded schools, and increases in K12 spending were more efficacious when preceded by Head Start exposure. The findings suggest dynamic complementarities, implying that early educational investments that are sustained may break the cycle of poverty. (*JEL* I20, J20)

* Jackson and Johnson 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).

I. Introduction

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, Klein, and Saez, 2014). 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, remediating the ill-effects of childhood poverty may require early investments in the skills of disadvantaged children that are followed by sustained investments over time.

This paper studies whether early childhood investments designed to promote school readiness among disadvantaged children that are followed up with increases in public school spending are particularly effective at improving their long-run outcomes. This question is one specific manifestation of the long-standing hypothesis in economics that, because skills beget skills, children who benefit from early human capital investments may benefit more from later investments (Cunha and Heckman, 2007). Testing this hypothesis is difficult, since it requires exogenous variation in multiple investments over time, a high bar that many previous papers have struggled to clear. Some earlier studies have examined whether the effect of human capital intervention varies by pre-intervention skill level (e.g. Garcia and Gallegos, 2017; Lubotsky and Kaestner, 2016; Aizer and Cunha, 2012). Because these studies do not use exogenous variation in prior skills, they do not speak directly to whether early and late human capital policies exhibit dynamic complementarity. Addressing this critique, other studies have examined whether the benefits of human capital investments vary among those who were exposed to non-investment skill shocks such as hurricanes, or rainfall during gestation (e.g. Adhvaryu, Molina, Nyshadham and Tomayo, 2017). Because these studies do not rely on human capital investments *per se* for variation in initial skills, they do not examine whether human capital investments made at different stages of the life course exhibit dynamic complementarity.

To test for dynamic complementarity we examine the interaction between two exogenous and independent human capital investment "shocks." The first exogenous shock to human capital investment is the rollout of Head Start, the largest early childhood intervention program in the US, which increased access to early childhood education and pediatric care for low-income children. The second exogenous shock to human capital investment is the implementation of court-ordered school finance reforms (SFRs) which (a) reduced differences in public K12 school spending between affluent and poor neighborhoods within states, and (b) increased (on average) the level of per-pupil spending at public K12 schools.[1] While dynamic complementarities may not exist between *any* two human capital investments, they are most likely to exist between two education-related interventions in which one (a school-readiness program) is designed to help children benefit from the other (public K-12 schools). Our setting is particularly well-suited for studying dynamic complementarities.[2]

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 2015 using the Panel Study of Income Dynamics (PSID). These data allow us to study potential complementarities on an array of adult outcomes including educational attainment, earnings, poverty, and incarceration. While test scores have been the traditional focus of evaluations of Head Start and K12 spending, the effects of interventions on long-run outcomes may go undetected by test scores.[3]

Identifying the interaction effects between two human capital investments requires that (a) one credibly identify the effects of each investment individually on the same individuals, and (b) that each of the human capital investments is independent of the other (Almond and Mazumder, 2013). To identify the causal effect of early childhood investments, we exploit variation in the timing of the rollout of Head Start across counties. 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. To identify the causal effects of public K12 school spending, we exploit geographic variation in the timing of

[1] See Card and Payne, 2002; Murray, Evans, and Schwab, 1998; Hoxby, 2001; Jackson, Johnson, and Persico, 2014 for a more complete disucssion of the effects of SFRs on public school spending.

[2] Investments are "broadly defined as actions specifically taken to promote learning" in Heckman and Mosso (2014). As such, interactions between education-related interventions are likely what the theory is about. In the only other paper to examine two education interventions, Gilraine (2016) finds that the benefits of accountability due to NCLB in later grades are larger among students exposure to accountability in earlier grades. Looking at a health and an education intervention, Rossin-Slater and Wust (2017) find that the effects of access to pre-school was *smaller* among those who had access to home visits during infancy. In work with a health intervention, Bhalotra and Venkataramani (2017) find that the benefits of antibiotic treatment for blacks decline in measures of the severity of institutionalized segregation. Also 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.

[3] E.g. Heckman, Pinto, and Savelyev, 2014; Jackson, forthcoming; Chetty et al., 2011; Ludwig and Miller, 2007.

court-ordered SFRs. Following Jackson, Johnson, and Persico (2016), (hereinafter JJP) 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 empirical tests to validate our models and to support a causal interpetation of the patterns presented.

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 the two. 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 the interaction effect to be identified, it requires that individuals that are exposed to both a SFR and Head Start are not somehow different from those that are exposed to only a SFR or only exposed to Head Start. One can only be confident that this condition is satisfied if both policy changes are independent of each other. We argue that because SFRs occurred at the state level (affecting all public schools in a state at the same time), while Head Start (a federal program) was introduced in certain counties within states at different times, these two policies are largely independent of each other. More formally, we show that (a) the raw correlation between the two policy instruments is only 0.15, (b) conditional on controls there is no association between Head Start spending and SFR-induced changes in K12 spending, and (c) using partial *F*-statistics, there is sufficient policy variation in Head Start spending and K12 spending for the effect of each to be identified and for the interaction between the two to be identified.

Our results show that both Head Start and SFR-based K12 spending increases have large, positive long-run effects, and we also find strong and robust evidence of dynamic complementarity. 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. We find no effect of Head Start spending on the outcomes of non-poor children. Increases in public school K12 spending improved this same array of outcomes in adulthood. Among poor children exposed to a 10% reduction in K12 spending, exposure to a typical Head Start center has small statistically insignificant effects on educational attainment, wages, incarceration, and adult poverty. However, among poor children exposed to a 10% increase in K12 spending, exposure to a typical Head Start center leads to 0.59 additional years of education, being 14.8 percentage points more likely to graduate high school, 17% higher wages, being 4.7 percentage points less likely to be incarcerated, and being 12 percentage points less likely to be poor as an adult.

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% 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, that same 10% increase in K12 per-pupil spending increases educational attainment by 0.4 years, increases earnings by 20.6 percent, and reduces the likelihood of incarceration by eight percentage points. The positive interaction effects between Head Start and K12 spending are robust across several models (including sibling comparisons) and are only present among poor children (who were eligible for Head Start). The effect of K12 spending was unrelated to the level of Head Start spending among non-poor children, for whom increasing K12 spending by 10% increased years of education by 0.2 and earnings by 11.7 percent.

The paper contributes to the literature in three ways. First, we provide direct evidence on the long-run benefits of both Head Start and K12 spending. Second, we present broad, robust evidence of complementarities between early and later human capital investments for low-income children. The complementarities imply that one could increase *both* equity and efficiency by redistributing spending from well-funded K12 schools toward Head Start programs targeted at poor children. Generally, our results are the first to show that early and sustained complementary

[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).

investments in the skills of low-income children can be a cost-effective strategy to break the cycle of poverty. Third, the use of quasi-experimental methods that involve two different, yet complementary, identification strategies yields a similar pattern of results and bolsters confidence in the overall set of findings.

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.

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). Informed by this research, Cunha and Heckman (2007), theorize that skill development is an interactive, multistage process in which the marginal effect of investments today is higher among those with a greater stock of previously acquired skills.[5] We refer to this characteristic of skills production as dynamic complementary *in skill development.* When this condition holds, "*skills produced at one stage raise the productivity of investment at subsequent stages"* (Heckman and Cunha 2007). We refer to such synergies between *investments* as dynamic complementary *in human capital investments*.[6] If Head Start increases skills and therefore improves school readiness, Head Start may facilitate better learning in the K12 system. If so, insofar as increased spending

[5] This is what is identified by researchers who examine the effect of interventions for individuals with differing incoming levels of skills (e.g. Garcia and Gallegos, 2017; Lubotsky and Kaestner, 2016; Aizer and Cunha, 2012).

[6] Following the notation from Heckman (2007), the technology of skills production is dynamic. Skills acquired when a child is *t* years old is [a] below

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

where *t*=1,2,...T, θ_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 human capital investments arises when the stocks of capabilities acquired by period *t-1* (θ_t) make investments in period t (I_t) more productive, i.e.,

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

Consider that $\theta_t = f_{t-1}(h_{t-1}, \theta_{t-1}, I_{t-1})$. Because $\partial f_t/\partial I_t > 0$, if [b] holds, then [c] below must also hold.

[c] $$(\partial^2\theta_{t+1})/(\partial I_{t-1} \partial I_t) > 0.$$

In words, dynamic complementarity in skill development implies that there is dynamic complementarity in human capital investments. However, if early investments increase the efficacy of later investments through mechanisms other than increasing skills, the converse does may not hold. We show that this is not the case in our setting.

improves school quality, spending on Head Start and public K12 schools would exhibit dynamic complementarity. This is what we seek to test in this paper.

Note that complementarity is not a given. Compensatory interventions or interventions designed to bring *all* children up to some basic standard of skill, may, by design, have smaller benefits for more highly-skilled children. Also, note that human capital investments may exhibit dynamic complementarity for reasons other than dynamic complementary in skill development.[7] To apply these insights to our setting, we outline two ways through which Head Start and K12 spending may interact. The first is a direct channel that operates through dynamic complementary in skill development. The second channel is indirect and may operate through spillovers to other students, and adjustments by actors in the schooling system (Malamud et al., 2017).

The direct channel operates through what we call "alignment." The alignment mechanism 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 require skills 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 alignment, 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 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,

[7] Formally, if equation [b] holds it implies that [c] will also hold. However, the converse is not always true.

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 first indirect channel is through "spillovers." Research has found that higher shares of low-performing peers or disruptive peers may have deleterious impacts on students (see Sacerdote, 2014). 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. This could make it easier for the K12 school system to translate resources into better outcomes.[8] The second indirect channel is through "adjustments." The first is an "alignment adjustment." 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), the quality of K12 instruction could be affected for all students. Importantly, because these adjustments can move some students closer to the target and others farther from it, the "alignment adjustment" effect could be positive or negative for any given student. There could also be "budget allocation adjustments" that can affect students in different classrooms. For example, lower shares of students requiring remediation or special services (due to Head Start) may allow schools to allocate resources to other productive inputs, which may affect all students in the school.[9]

This is not an exhaustive list of all possible adjustment effects. However, the key takeaways are that (a) policy complementarities reflect both the direct effect due to the technology of skill formation and also some spillover and adjustment effects, and (b) adjustment and spillover effects could lead the interaction between the two interventions to be either positive or negative such that the overall interaction effect is ambiguous in sign. We present empirical evidence to shed light on what mechanisms are most likely at play in our setting.

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

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

[8] 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.

[9] Head Start also teaches parenting skills; thus, another possible indirect channel is changes in parental quality. We test for this using within-family variation. We find no indication that siblings of those exposed to Head Start have improved outcomes (Appendix I). This runs counter to the parental quality mechanism.

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 nutritious meals for the children. Participating children receive development screenings, and programs connect families with medical, dental, and mental health services. 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. 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).

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.[10] To open a new Head Start center, local organizations (typically non-profit organizations, for-profit agencies, or school systems) apply to the federal government for grant funds. Grantees provide at least 20% of the funding. After approval, Head Start grants are awarded directly to applying organizations subject to three-year grant cycles. 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). 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).

Head Start was targeted at pre-school age children (3 through 5) and most Head Start enrollees were four years old at enrollment. 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. The top panel of Figure 1 plots the raw national Head Start enrollments between 1960 and 1994. 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

[10] 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.

Start. As such, the early rollout of Head Start represented *both* increases in Head Start participation *and* enhancements in the Head Start programs themselves. Another notable pattern is the decline in Head Start enrollments between 1969 and 1972. 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). To relate these enrollments to participation rates at the individual child level, for each kindergarten entry cohort we computed the cumulative likelihood across all age-eligible years that an income-eligible child would enroll in Head Start.[11] The lower panel of Figure 1 depicts our estimated likelihood of Head Start enrollment (across all age-eligible years) by kindergarten entry cohort. The likelihood of Head Start enrollment among poor children reached 86% for income-eligible cohorts entering kindergarten in 1969, fell in the early 1970s, and stabilized around 63% by 1990. This is similar to the Garces et al (2002) estimate of two-thirds. Our participation rates of between 63 and 85 percent are important to keep in mind as we interpret the magnitudes of our intent-to-treat estimates (in Section VI). Figure 1 (top panel) also plots the share of 3- and 4-year-olds enrolled in full-time daycare over time (as reported in the Current Population Survey). 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 ages 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.

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.[12] Because the target population for Head Start is poor pre-schoolers and most enrolles are 4 years old, 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

[11] The ratio of enrolled students to the income-eligible age-eligible population in a given year *is not* the same as a specific cohort's participation rate by kindergarten entry. See Appendix M for a more detailed discussion of this. To avoid double-counting individuals who enrolled in both the summer program and the full-year programs, we assume that 40 percent of full-year enrollees were previously in a summer program.

[12] See Appendix B for an illustration discussion of this.

most counties, the *first* Head Start center was established between 1965 and 1970.[13] 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.

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.[14] 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.

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. We follow JJP and categorize reforms into four types. Foundation plans guarantee a base level of per-pupil spending and increase per-pupil spending for the lowest-spending districts. Spending-limit plans prohibit per-pupil spending levels above some predetermined amount. Reward-for-effort plans match locally-raised funds for education with additional state funds (often with higher match rates for lower-income areas). Equalization plans typically tax all districts and redistribute funds to lower-wealth and lower-income districts. These reform/formula types are not mutually exclusive.

In existing work Card and Payne (2002), JJP and Hoxby (2000) find that court-ordered SFRs that lead to the implementation of different funding formulas had different effects on district

[13] Figure A2 presents each county in the United States color-coded by the year of its first Head Start center.
[14] See Jackson, Johnson and Persico (2016) for a full discussion of SFRs.

spending by pre-reform income and spending levels. [15] In particular, JJP find that reforms that lead to "reward-for-effort" formulas tended to increase per-pupil K12 spending in all districts; spending limits led to 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 increases in each district as a function of the reform type introduced (by the state) and the pre-reform characteristics of the district. Because these relationships are unrelated to 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.[16] 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 are linked to a database of SFRs from JJP.[17] 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, 1968-2015), and our analysis sample includes individuals born between 1950 and 1976 who were 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. We include all information on PSID individuals between 1968 and 2015.[18] We linked

[15] To illustrate how the introduction of different formula types affected districts by pre-reform income and spending levels, we replicate the analisis in Jackson Johnson and Persico (2016). This is in Appendix C.
[16] Further details on the PSID data are in Appendix D.
[17] Details on how these databases were compiled and the coverage of districts in these data are in Appendix E.
[18] 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. The PSID maintains high wave-to-wave response

persons in the PSID using their census blocks during childhood to school spending data, SFR data, and Head Start spending data. 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 detail the algorithm in Appendix D. Among potentially treated cohorts, 97 percent of the earliest address information is from *before* the policies we study were enacted so that bias due to residential sorting in response to the policies is negligible. We verify this empirically. 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.[19]

We define low-income children as those whose average parental income (between ages 12 and 17) fell in the bottom quartile.[20] Among cohorts born between 1963-1976 for whom parental income at age four is observed, roughly 80% of those whom we classify as low-income were below the federal poverty line at age four, and 93% of those who were below the poverty threshold at age four are classified as low-income by our definition. The analytic sample includes 15,232 individuals from 4,990 childhood families, 1,427 school districts, 1,120 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[21] (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

---

rates of 95-98%. We perform a supplementary analysis of sample attrition in the PSID, and find no evidence of selective attrition among our study sample (Appendix Table D1).

[19] See Appendices D and E for a discussion of these data sources.

[20] 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.

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

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). We take advantage of the staggered introduction across geographic areas of Head Start programs during the program's rollout. Before the rollout of Head Start to an area, there is no Head Start spending. After the introduction of Head Start in a county, spending levels subsequently increase. Figure 2 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.

Once the first center is established, spending per poor four-year-old increases more rapidly in the high- than the low-spending counties (left panel). 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 and reverted to near zero Head Start spending within four years. 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. Figure 2 reveals exactly this pattern for years of educational attainment (measured in adulthood) among poor children. Areas with small (middle panel) and large increases in Head Start spending (right panel) were on similar trajectories among cohorts who were older than four years old when the first Head Start center was established (i.e., years -5 through year 0). However, the post-rollout cohorts have much better outcomes in high Head Start spending counties than in low-spending counties.

Our preferred difference-in-difference (DiD) strategy uses 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 [1] by Ordinary Least Squares (OLS).

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

In [1], 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, [1] includes childhood county fixed effects (θ_c); and to account for cohort effects we include birth-year fixed effects (τ_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 ε_{icb}.

There are two identifying assumptions. First, counties that experienced larger or smaller increases in Head Start spending over time 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 2 suggests that the first condition is satisfied. To show that the second assumption is likely satisfied, we examine whether areas that had higher levels of Head Start spending may have also introduced policies and programs that may have improved child outcomes.

To test this, we estimated the marginal effect of Head Start spending levels that prevailed when individuals were different ages, conditional on the level of Head Start spending when they were four (shown in the left panel of Figure 3 and Table H8). 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.[22] If areas with high levels of Head Start spending also implemented other policies that would promote better outcomes, then Head Start spending at age 5, 3, or 7 would systematically be associated with better outcomes. This is clearly not the case. As another check on this variation, we regress each person's years of education and wage on our rich set of individual, family, and neighborhood characteristics and other social safety net programs. The fitted values from these regressions are effect-size weighted indices of childhood family and community socioeconomic factors (Appendix Table H3). Conditional on school-district and birth-year fixed effects only, there is no association between Head Start spending and these predicted outcomes. Taken together, this is compelling evidence that our variation is valid. However, to assuage any lingering worries, we also implement a second strategy.

Because local areas with high versus low levels of Head Start spending may differ in ways that could confound our comparisons, our second identification strategy relies only on the variation

[22] Appendix Figure H1is an analogous figure for adult wages.

in the availability of any local Head Start center at age four. To do this, we instrument for Head Start spending per poor four-year-old in county c ($HS_{cb}^{age\ 4}$), with an indicator variable of whether a Head Start center existed in one's childhood county at four years old ($Exposed\_HS_{cb}^{age\ 4}$). Formally, we estimate the following system of equations by two-stage-least-squares (2SLS)

$$[2] \qquad \widehat{HS}_{cb}^{age\ 4} = \pi_{hs,1} \cdot Exposed\_HS_{cb}^{age\ 4} + \pi_{hs,1} \cdot C_{icb} + \theta_c + \tau_b.$$

$$[3] \qquad Y_{icb} = \beta^{2SLS} \cdot \widehat{HS}_{cb}^{age\ 4} + \gamma \cdot C_{icb} + \theta_c + \tau_b + \varepsilon_{icb}.$$

The identifying assumptions in this 2SLS model are weaker than those for the DiD model. This model is identified if (a) counties that establish a Head Start center were not already on a trajectory of improving or deteriorating outcomes over time, and (b) counties that established a Head Start center did not also undergo other unobserved changes that would also affect outcomes. The right panel of Figure 3 (and Table H9) shows the effect of rollout (as opposed to spending) by age on years of education for poor children.[23] Reassuringly, there is an effect of having access to Head Start at age 4 but no effect of having access to Head Start for any other age (conditional on access at age 4). Figures 2 and 3 suggest that the identifying conditions are satisfied. Furthermore, in Section VI.B we present further evidence to support a causal interpretation of our estimates.

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.[24] 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 (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

---

[23] Appendix Figure H1is an analogous figure for adult wages.
[24] We use the natural log to capture the fact that school spending likely exhibits diminishing marginal product.

than others. We exploit this fact and 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 JJP, 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 [4] by 2SLS. All common variables are defined as in [1].

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

To rely only on variation across birth cohorts within districts, we include school district fixed effects (θ_d); to account for time trends and cohort effects, we include birth-year fixed effects (τ_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, $SFRExp_{idb}$, 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 [5] below

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

Following JJP, $\widehat{dose}_d$, is a predicted reform-induced spending change for each district based on reform type (implemented at the state level), pre-reform district income levels, pre-reform district spending levels and their interactions.[25] By construction, $\widehat{dose}_d$ is unrelated to endogenous decisions made by districts after reforms. Because we estimate $\widehat{dose}_d$ using *all* school districts while we estimate effects using the PSID sample, our approach is a two-sample-2SLS.[26] To assuage any concerns regarding $\widehat{dose}_d$, Table H2 shows that the estimated point estimates obtained when using only variation in SFR exposure are almost identical (albeit less precise) than those that use both exposure and exposure times dosage.

Figure 4 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. $\widehat{dose}_d \leq 0$).[27] 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. Accordingly, negative values are cohorts who were 18 or older at the passage of a court-ordered SFR, the "0" cohort was 17 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 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. As shown in JJP, unexposed cohorts in districts with lower and higher predicted dosage were on similar pre-reform trajectories; however, exposed cohorts in high dose states experienced much larger increases in per-pupil spending after a SFR. This shows that 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-

[25] 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. This regression models how per-pupil spending evolves in a district after the passage of a court-ordered SFR as a function of the funding formula introduced in the state, the school spending level in the district, and the economic characteristics of the district *prior* to reforms. We take the fitted values from this regression to obtain a predicted reform-induced spending change for each district (based on these exogenous variables). See Appendix F and Jackson Johnson and Persico (2016) for more details.

[26] 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).

[27] 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 4, 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.

reform district characteristics) isolates exogenous variation in school spending.

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 4 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.

The key identifying assumptions are that (a) districts that experienced spending increases due to a SFR were not on different trajectories before reforms, and (b) there were no coincident district-level policies or changes that confound our analysis. Figure 4 shows that this first condition is likely satisfied. We also test the second condition. If other coincident policies were driving the results (that were not targeted to school-age children), increased school spending might improve outcomes of those who were in the same district but not of school-going age. To test this, we instrument for the K12 spending levels that prevailed in an individual's childhood district when they were between the ages of 18 to 22 (i.e., non-school-going age). We find no effect on adult outcomes (Appendix Table H1). Also, we find that conditional on school-district and birth year fixed effects, there is no association between instrumented K12 spending and predicted outcomes (Appendix Table H3) – further evidence that our identifying variation is valid. While these tests are not dispositive, they support a causal interpretation of the main findings. To assuage any lingering concerns, we present additional tests in Section VI.

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 $\mathrm{INT}_{idb} = \left(HS_{icb}^{age\ 4} \times ppe_{idb}^{5-17}\right)$. We estimate two different models in our analysis.

The DiD-by-2SLS model. In the first model we use the within-county, across-cohort DiD variation in Head Start spending ($HS_{icb}^{age\,4}$). Because a school district may be a smaller unit of observation than a county, all models include district fixed effects (which subsumes county effects). We instrument for ppe_{idb}^{5-17}, with ($SFRExp_{idb}$) and ($SFRExp_{idb} \times \widehat{dose}_d$). We instrument for, INT$_{idb}$ with ($HS_{icb}^{age\,4} \times SFRExp_{idb} \times \widehat{dose}_d$) and ($HS_{icb}^{age\,4} \times SFRExp_{idb}$).[28] The resulting model is [6], where $\widehat{ppe}_{idb}^{5-17}$ and $\widehat{INT}_{idb}$ are fitted values from first-stage regressions.[29]

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

The 2SLS-by-2SLS model. In the second model, we instrument for *all* spending variables. Now we instrument for Head Start spending ($HS_{icb}^{age\,4}$) using exposure to any Head Start center at age 4 ($Exposed\_HS_{cb}^{age\,4}$). We instrument for ppe_{idb}^{5-17}, with ($SFRExp_{idb}$) and ($SFRExp_{idb} \times \widehat{dose}_d$). Now we instrument for, INT$_{idb}$ with ($Exposed\_HS_{cb}^{age\,4} \times SFRExp_{idb} \times \widehat{dose}_d$) and ($Exposed\_HS_{cb}^{age\,4} \times SFRExp_{idb}$). The resulting model is as in [7], where $\widehat{ppe}_{idb}^{5-17}$ and $\widehat{INT}_{idb}$ and $\widehat{HS}_{cb}^{age\,4}$ are *all* fitted values from first-stage regressions. [30]

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

The interaction effect between pre-K and K12 spending can be identified because (a) among counties that faced similar increases in Head Start spending (or had any Head Start center), 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 no) 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

---

[28] While intuition would lead one to expect us to use all the two-way interactions between $HS_{icb}^{age\,4}$, $\widehat{dose}_d$, and $SFRExp_{idb}$, we do not use ($HS_{icb}^{age\,4} \times \widehat{dose}_d$) as an excluded instrument because $\widehat{dose}_d$ cannot affect outcomes unless it is interacted with SFR exposure. This would simply introduce noise and weaken the first stage.

[29] Where $\hat{X}_1 = \widehat{ppe}_{idb}^{5-17}$ and $\hat{X}_2 = \widehat{INT}_{idb}$, and $w \in \{1,2\}$,
$\hat{X}_w = \pi_{w1}(SFRExp_{idb} \times dose_c) + \pi_{w2}(SFRExp_{idb}) + \pi_{w3}\left(SFRExp_{idb} \times \widehat{dose}_d\right) \cdot HS_{cb}^{age\,4} + \pi_{w4}(SFRExp_{idb}) \cdot HS_{cb}^{age\,4} + \gamma_w C_{idb} + \theta_{wd} + \tau_{wb}$.

[30] Where $\hat{X}_1 = \widehat{ppe}_{idb}^{5-17}$, $\hat{X}_2 = \widehat{INT}_{idb}$, $\hat{X}_3 = \widehat{HS}_{cb}^{age\,4}$ and $g \in \{1,2,3\}$,
$\hat{X}_g = \pi_{g1}(SFRExp_{idb} \times dose_c) + \pi_{g2}(SFRExp_{idb}) + \pi_{g3}\left(SFRExp_{idb} \times dose_c \times Exposed\_HS_{cb}^{age\,4}\right) + \pi_{g4}\left(SFRExp_{idb} \times Exposed\_HS_{cb}^{age\,4}\right) + \gamma_g C_{idb} + \theta_{gd} + \tau_{gb}$.

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 childhood county-by-birthyear 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.

To provide visual evidence of complementarity, Figure 5 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. This is the variation used in the 2SLS-by-2SLS models. The left panel shows that SFR-treated and 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 Head Start 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 poor children. However, if the two policies are complementary, one should see similar patterns and greater improvements for large increases in K12 spending. This is precisely what we document in the right panel of Figure 5. In districts that experienced large increases in K12 spending after a SFR, exposed cohorts achieve more years of education than unexposed cohorts, and the relative increase is larger among those SFR-exposed cohorts that were from counties with a Head Start center at age four. Furthermore, the benefits of Head Start spending (the difference between the grey and black line in each panel) are larger among SFR-exposed cohorts that experience larger K12 spending increases. In sum, Figure 5 presents visual evidence that Head Start and K12 school spending exhibit dynamic complementarity.[31] The lack of any differential pre-trending in either panel illustrates that the parallel trends assumption likely holds, not just for each policy (Figures 2 through 4), but also for the *interaction* between the two policies.[32]

V.C.1. *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. For our inference to be valid, these policy instruments need to be largely independent of each other. This is necessary for two reasons. First, if there were a high correlation between the two policy instruments, a model predicting both the base effects and the interaction could be under-identified. If so, there would be a weak first stage for the interaction, conditional on the instruments for the base effects. Second, if areas that were most likely to have high levels of Head Start spending were also likely to have larger SFR induced K12 spending increases, then areas that were exposed to high levels of both may differ from areas that were only exposed to only one, or none in unobserved ways. Because our interaction is essentially a comparison of Local Average Treatment Effects (LATEs), ***if*** there is treatment heterogeneity, the resulting interaction effect ***could*** simply reflect a difference in LATEs rather than a true interaction effect.

We show that this is not a problem in our setting. First, the correlation between Head Start spending and instrumented K12 spending is only 0.15, and conditional on controls, there is no association between the two (Appendix Table H6). This suggests that our treatments are largely independent so that we are not comparing different LATEs. Also, following Angrist and Pischke (2009), we compute first-stage F-statistics for each set of excluded instruments, conditional on the other excluded instruments. The first-stage F-statistic on the instruments for K12 spending (i.e., predicted SFR dosage times years of SFR exposure) is 22.41 and 23.01 in models without and with Head Start variables included, respectively (Appendix Table H7). The first-stage F-statistic on the excluded instruments for Head Start spending (i.e., the existence of a Head Start center at age 4) is 59.17 and 60.76 in models without and with the K12 instruments included, respectively. Finally, the first-stage F-statistic on Head Start Exposure times SFR dosage times SFR exposure is 42.46,

[31] An analogous figure for adult wages is in Figure L1. Though muted, one can see the same basic patterns.

[32] If the alignment channel is at play, complementarity would be larger for cohorts that were exposed to a SFR soon after Head Start than for those exposed later. The patterns in Figure 5 are consistent with this. Specifically, the cohorts that are exposed to an SFR for more years benefit much more from Head Start than those exposed to a SFR for fewer years (e.g., 8-10 years vs 2-6 years). To present another suggestive test, we estimate our main models and interact all of our K12 spending variables with indicators measuring the age at which the SFR was implemented in one's childhood state. The results (Appendix K) suggest that the complementarity effects are driven by those cohorts that were exposed to an SFR before the age of 9.

conditional on Head Start Exposure, SFR exposure, and SFR dosage times SFR exposure. In sum, there is sufficient 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. RESULTS

We present results from specification [6] that exploits all the within-district, across-cohort variation in Head Start spending and instruments for K12 public school spending using the SFR instruments, and specification [7] that instruments for both Head Start spending and K12 spending. 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 for poor children. Column 1 presents the DiD-2SLS estimates of the effects on the probability of graduating from high school. The coefficient on Head Start spending per poor four-year-old is 0.025 (p-value<0.01). That is, increasing Head Start spending per poor 4-year-old in the county by $1,000 (roughly a 25% increase) increases the likelihood of graduating from high school by 2.5 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 $4,230, this implies that, for poor children, having access to the average Head Start program increased the likelihood of graduating from high school by roughly 10 percentage points. Column 2 presents effects for the 2SLS-2SLS design that instruments for all spending variables. The 2SLS coefficient on Head Start spending per poor four-year-old is quite similar (it is 0.0408 and is statistically significant), and one cannot reject that the DiD-2SLS models and the 2SLS-2SLS models yield different results. However, in the fully instrumented model, the effect of Head Start spending is slightly *larger* and less precisely estimated. Because the DiD-2SLS estimated Head Start effects tend to be smaller, and the results are similar to the fully instrumented model, we take a conservative approach and focus discussion on the DiD-2SLS results. However, Table 2 reports all results from the 2SLS-2SLS models.

Increases in Head Start spending can affect outcomes through increases in 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.

We estimate that the rollout of Head Start increased Head Start participation for poor children by about 75 percentage points. We come to this conclusion in two ways.[33] First, using national data, for cohorts entering kindergarten after 1966, the likelihood of Head Start enrollment (full-time or part-year) among income-eligible children was 63% (Figure 1). Because centers can enroll 10% of non-poor children, the participation rate among income-eligible children *could have been* as low as 57 percent. Roughly 80% 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 0.57/0.8=0.71 (i.e. 71 percentage points), conditional on having a Head Start center in the county at age 4. We arrive at a similar estimate using retrospective survey questions from the PSID. In the 1995 survey wave as part of a special module on early childhood, adults were asked about whether they had ever participated in a Head Start program. These data may have a number of limitations such as recall bias (See Appendix G). However, in these data, Head Start rollout increases Head Start participation among poor children by about 80 percentage points. We use 75 percentage points as our "ballpark" estimate of the increase in the likelihood of Head Start participation (among poor children) due to the rollout of the average Head Start center in the county during our study period.

If all of our estimated effect of having Head Start access was due to Head Start enrollment

[33] The PSID survey data employed in Garces, Currie, and Thomas (2002) are retrospective data collected in the 1995 wave. There are some concerns about potential measurement error and recall bias in using this retrospective survey information about Head Start participation and some missing information. See Appendix G for further discussion.

(and there were no spillover effects to other poor children), our participation margin effect implies a treatment-on-the-treated effect of 0.1/.75=0.129, or 13.3 percentage points. This is similar to the estimated enrollment effect of Head Start in existing studies.[34] However, most existing studies of Head Start focus on full-year Head Start programs. If one makes the conservative assumption that there is no effect of summer-only programs or part-time programs, a back-of-the-envelope calculation yields an implied treatment-on-the-treated effect of full-year Head Start on the likelihood of high school graduation of 15.3 percentage points.[35] This estimate is in line with the larger of the participation margin effects in the 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 poor non-participants in the K12 school system.

Our K12 spending results replicate JJP. The coefficient on the log of K12 spending during the school-age years is 1.10 (p-value<0.01). Increasing K12 school spending (across all 12 school-age years) by 10% increases the likelihood of high school graduation by about 11 percentage points for a poor child exposed to the average level of Head Start spending (Column 1). Relative to baseline, this is about a 15% increase. The estimates indicate that increasing Head Start spending by $4,000 would have roughly the same effect on high school graduation as increasing K12 spending by 10% across all school-age years (for poor children).[36]

Columns 3 and 4 present a similar pattern for completed years of education for poor children. The more conservative DiD-2SLS estimates reveal that increasing Head Start spending per poor 4-year old in the county by $1,000 increases the years of educational attainment by 0.077 years (p-value<0.01) for a poor child exposed to the average level of K12 spending. At average Head Start spending levels, a Head Start center is estimated to increase years of education by

[34] For example, Garces, Currie, and Thomas (2002) 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.
[35] The average enrollment rate among eligible children was 52% after the initial ramp up period (for cohorts entering kindergarten after 1966). This implies a full-year Head Start participation rate of about 0.52/0.8=0.65 conditional on having a Head Start center in the county at age 4. If one makes the conservative assumption that there is no effect of summer only programs or part-time programs so that *all* of our estimated intention-to-treat effect was due to *full-year* Head Start enrollment, an assumed upper-bound full-year Head Start participation margin effect implies a treatment-on-the-treated effect on the likelihood of high school graduation of 0.1/.65=0.153.
[36] During the sample period, a 10% 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% interest rate).

roughly a third of a year. Increasing school-age K12 spending by 10% increases the number of years of completed education by about 0.4 years for a poor child exposed to the average level of Head Start spending. As discussed previously, the 2SLS-2SLS results in column 4 are similar.

Results for non-poor children are in Table 3. As in JJP, the estimated K12 spending effects on the education outcomes are positive, sizable, and statistically significantly different from zero. This indicates that increases in K12 spending improve the educational outcomes of not only the poor but also the non-poor. The DiD-2SLS point estimates indicate that increasing district K12 spending by 10% increases the likelihood of graduating high school by 2.3 percentage points, and increases years of educational attainment by about 0.24 years for a non-poor child exposed to the average level of Head Start spending. These estimated K12 spending benefits are smaller for more affluent children than for poor children, but they are positive, statistically significant, and economically important. In contrast to the positive K12 spending effects, for non-poor children, both the DID-2SLS and 2SLS-2SLS model reveal that increasing Head Start spending has small, insignificant effects. For both education outcomes, one cannot reject that the effect on the non-poor is zero, and one *can* reject that the Head Start effect is the same for both poor and non-poor children.[37] 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 broad policies that improve the outcomes of non-poor children.

The fact that we find no effect of Head Start spending for non-poor children is important. If local areas that increased Head Start spending introduced other policies that improve outcomes of all children, one would observe positive Head Start spending effects for the non-poor children. We find no such pattern. Our result, instead, implies that neither our variation in Head Start spending nor the rollout of Head Start is associated with any policies that improved the outcomes of local children who were ineligible to participate in Head Start. This coupled with the fact that Head Start spending only influences outcomes for those who were four years old at the time shows that we only see effects for children who were both income- and age-eligible for Head Start. This serves as another falsification test, of sorts, and bolsters the credibility of the research design.

The adult economic outcomes we examine are wages and the annual incidence of poverty

[37] We pooled the samples and estimated a single model where we interacted all variables with poverty status, and tested for equality of coefficients between poor and non-poor children for our key explanatory variables. We present the results of this test for our two main adult outcomes in the conservative DiD-2SLS models in Appendix Table J1. Our tests reject that the estimates are the same for the two populations.

between the ages of 20 and 50. Our models 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 5 through 8 in Table 2 present these results for children from poor families. Looking at wages, in the DiD-2SLS models (column 5) the coefficient on the log of public K12 school spending is 2.056 (p-value<0.1) and that on Head Start spending per poor 4-year-old is 0.023 (p-value<0.01). That is, for children from poor families exposed to average levels of Head Start spending, increasing K12 spending by 10% is associated with about 20.5% higher adult wages. Similarly, for these same children, at average public K12 spending levels, increasing Head Start spending by $4,230 per poor 4-year-old (the average spending amount) is associated with 9.87% higher wages for poor children. The results in the 2SLS-2SLS models are similar and cannot be distinguished statistically.

Columns 5 and 6 of Table 3 present the effects on adult wages for non-poor children. Similar to the educational outcomes, there are positive effects of K12 spending, but no effect of Head Start spending on the wages in adulthood of those from non-poor families. In the DiD-2SLS models, the coefficient on the log of K12 public school spending is 0.7351 (p-value<0.05), and that on Head Start spending per poor 4-year-old is 0.0069 (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% is associated with 7.35% higher earnings between the ages of 20 and 50, while increasing Head Start spending is associated with no difference in earnings. The 2SLS-2SLS results (column 6) tell the same basic story as the DiD-2SLS models.

The pattern of estimates for the annual incidence of poverty in adulthood in columns 7 and 8 of Tables 2 and 3 mirror those for adult wages. 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 and wage, it is a measure of hardship. Among poor children, Head Start spending is associated with large, statistically significant reductions in the annual incidence of poverty in adulthood (Table 2); while Head Start has small, insignificant effects on the adult outcomes of non-poor children (Table 3). However, increases in public K12 spending are associated with significant reductions in the likelihood of poverty in adulthood for all children, on average.

The final outcome we examine is the probability that an individual has ever been incarcerated (Column 9 and 10 of Tables 2 and 3). In the DiD-2SLS model, for poor children (Table 2), a $1,000 increase in Head Start spending reduces the likelihood of being incarcerated by 0.6 percentage points (p-value<0.01). This implies an average Head Start rollout effect (i.e., an increase of $4,320) of 2.5 percentage-points lower likelihood of adult incarceration (at average public K12 spending level). If one were to ascribe all of this effect to the participation margin for full-year Head Start, it would imply a Head Start participation effect of a five-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 9 also shows that increasing K12 per-pupil spending by 10% (at average Head Start spending levels) reduces the likelihood of adult incarceration by eight 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. The 2SLS-2SLS models in column 10 yield similar patterns, but with somewhat larger Head Start effects and wider confidence intervals. Looking at non-poor children (Table 3), we find no effect of either Head Start or K12 spending on the likelihood of adult incarceration among non-poor children. We attribute this to the low levels of incarceration among non-poor children. Importantly, *as with the other outcomes*, Head Start spending has no impact on those who were not income-eligible to participate.

VI.B. *Testing for Bias due to Unobserved Family Differences*

While we have presented much evidence that our variation is exogenous to other policies that may have been implemented in a locality, we have not *yet* ruled out the possibility that our results are driven by unobserved differences across treated and untreated families within local areas. To do this, 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 [6] and [7] to include sibling fixed effects (see Appendix Table H4). In such models, effects are similar to those in Table 2 so that unobserved family differences cannot explain the main pattern of results.

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 [6] and [7], we test whether the coefficient on the interaction is positive and statistically significantly different from zero. Across all outcomes for poor children, and across all specifications, increases in Head Start spending raise the marginal effect of K12 spending and *vice versa* – that is, all of the interaction terms are statistically significant at, *at least*, the 10 percent level in all models for all outcomes (Table 2). 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 statistically significant, and the signs of the coefficients across outcomes do not go in the same direction.[38] That is, Head Start spending had no direct or indirect effect on the outcomes of non-poor children.

To show the impact of these interaction effects, we present the marginal effects of each intervention evaluated at different levels of the other. Specifically, using the regression estimates, we compute the marginal effect of increasing Head Start spending per poor four-year-old by $4,230 when there is a 10% decrease, no increase, and a 10% increase in K12 spending (conditional on the direct effect of the change in K12 spending). Similarly, we compute the marginal effect of increasing K12 spending by 10% where there is no Head Start in the county and counties with average Head Start spending ($4,230). The estimated marginal effects for each model is presented in the lower two panels of Tables 2 and 3. As before, we focus on the DiD-2SLS models, but we present the 2SLS-2SLS models to show robustness.

Looking at high school graduation among poor children, having a Head Start center with a 10% decrease in K12 spending increases high school going by a statistically insignificant 6.3 percentage points. However, having a Head Start center with a 10% increase in K12 spending increases high school going by a 14.87 percentage points (p-value<0.01). The marginal effect of Head Start is more than twice as large when followed by a 10% increase in K12 spending than when followed by a 10% decrease. Also, the marginal effect of Head Start when there is a 10% decrease in K12 spending (though economically meaningful) cannot be distinguished from zero is a statistical sense. These patterns hold in both DiD-2SLS and 2SLS-2SLS models.

We now quantify the interaction concerning the marginal effect of K12 spending. The DiD-2SLS results indicate that increasing K12 spending across all school-age years by 10% increases the likelihood of graduating high school by 6.7 and 11 percentage points, with and without Head Start, respectively. Similar comparisons for children from non-poor families reveal that the effect of K12 spending on the outcomes of the non-poor is similar irrespective of the level of Head Start,

[38] We formally test that the marginal effects of Head Start and the "HeadStart*K12" interaction are different for poor children and non-poor children for years of education and adult wages. We do this by stacking the data and testing for equality of the coefficients. We present this test for years of education and wages in Appendix K.

and Head Start has no effect on the outcomes of the non-poor irrespective of the level of K12 spending. Because the DiD-2SLS results are similar to, but more conservative than the 2SLS-2SLS estimates, we focus on these models for the remainder of the paper.

The pattern of results for years of completed education is similar to those for high school graduation. The DiD-2SLS results are presented graphically in Figure 6 (the underlying estimates are in Table 2 and 3). For poor children (left panel), access to the average Head Start center increases completed education by 0.0533 years with a 10% reduction in K12 spending, increases education by 0.32 years with no change in K12 spending, and increases education by 0.599 years with a 10% increase in K12 spending. While the effect of Head Start with a reduction in K12 spending cannot be distinguished from zero, the effect when coupled with a 10% increase in K12 spending is statistically significant at the 1% level. For non-poor children (right panel), there is no effect of Head Start irrespective of the increase in K12 spending. Looking at the effect of K12 spending, for poor children, increasing K12 spending by 10% increase the years of education by 0.13 and 0.4 years, without and with Head Start, respectively. The effect of K12 spending is more than twice as large among poor individual exposed to Head Start than those who are not. For children from non-poor families (who are not eligible for Head Start), increasing K12 spending by 10% lead to about a 0.23 more years of education irrespective of the Head Start exposure.

In sum, these patterns suggest important dynamic complementarity between early childhood education spending and public K12 spending for the educational outcomes of poor children. In fact, due to the dynamic complementary for poor children, the pattern of results indicate that in areas with Head Start programs, increases in K12 spending both increased outcomes for all students and simultaneously reduced educational attainment gaps. The fact that there is no evidence of complementarity for non-poor children is important. It suggests that our main effects are not simply picking up some strange LATE for those places that happen to be exposed to both high K12 spending levels and Head Start. If our effects were due to this, one would observe positive interaction effects for all children in such districts. Instead, we find no interaction effects for the non-poor – indicating that our diagnostic tests were likely valid and further supports that our empirical strategy credibly identifies the interaction effects.

Commensurate with the educational outcomes, there is evidence of complementarity between Head Start spending and public K12 spending in the production of adult economic outcomes for children from poor families. Because for non-poor children there are no interaction effects for any outcome, we focus on the results for poor children. Figure 7 presents the marginal effect on adult wages of K12 spending by Head Start access (and vice-versa). For poor children (left panel), access to Head Start (with average funding levels) increases adult wages by 2.7% (p-value>0.1) when coupled with a 10% K12 spending decrease, increases it by 9.8% when there is no change in K12 spending (*p*-value<0.01), and increase wages by 17% when coupled with a 10% increase in K12 spending (*p*-value<0.01). The dynamic complementarities are sufficiently large that the marginal effect of the same increases in Head Start spending on the adult wage is about 70% larger when K12 spending increases by 10% than with no change. Looking at the effects of K12 spending increases, a 10% increase in K12 spending leads to 13% higher wages without Head Start, and 20% higher wages with Head Start (both effects are significant at the 1% level).

The effects on the annual incidence of adult poverty are consistent with those on education and wages (Columns 7 and 8 of Table 2). For poor children, increasing Head Start spending from zero to average levels reduces the annual incidence of poverty in adulthood by about 3 percentage points (p-value>0.1) when coupled with a 10 reduction in K12 spending, a 7.6 percentage point reduction when coupled with no change in K12 spending (*p*-value<0.01), and reduces adult poverty by 12 percentage-points when coupled with a 10% increase in K2 spending (*p*-value<0.01). The marginal effects of K12 spending tell the same story. A 10% increase in K12 spending leads to 3.3 and 7.96 percentage-points lower adult poverty without and with Head Start, respectively. The effect of the K12 spending increase with Head Start is significant at the 1% level and is more than twice as large as the effect with no Head Start.

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*. The marginal effects are presented in (Columns 9 and 10 of Table 2). For poor children, increasing Head Start spending from zero to average levels has no effect on the likelihood of incarceration when coupled with a 10% reduction in K12 spending. However, this same increase in Head Start exposure leads to a 2.5 percentage point reduction when coupled with no change in K12 spending (*p*-value<0.1), and a 4.73 percentage point reduction when coupled with a 10% increase in K2 spending (*p*-value<0.01). Looking to the effect of K12 spending on the likelihood of being incarcerated, the marginal effects are larger with Head Start than without. A 10% increase in K12 per-pupil spending reduces the likelihood of being incarcerated by 5.8 percentage points with no Head Start spending (*p*-

value<0.05), and by eight percentage points with Head Start (p-value<0.01).

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

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 with the same parents, we can test for improvements in parenting quality. We 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 (a) parenting quality is not part of the story, (b) our Head Start spending effects reflect real investments in the human capital of poor children and (c) our effects are not due to other confounding policies aimed at poor 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 [6] 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. 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). In any given location, 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. Using the esitmated impacts from Tables 2 and 3, we compute the marginal impact on the average outomes in a county of establishing an typcial Head Start center that only poor children beenfit from π_{HS}.

We also compute the marginal impact of spending that same amount of money (in present value terms) in the K12 system that all students attend π_{K12}. We then compute the ratio of these marginal impacts, π_{HS}/π_{K12}, for different poverty levels p. See Appendix N for details of this calucation. When spending is allocated optimally, this ratio should be 1.

In Figure 8 we plot this ratio against the poverty rate, where this ratio is evaluated at the mean level of K12 (i.e., using the empirical estimates from Table 2). We show this for adult wages (effects are similar for other outcomes). Because our empirical model is linear in Head Start spending but linear in the *log* of K12 spending, the marginal effect of K12 spending will fall relative to that for Head Start at higher levels of K12 spending even without any complementarity. To show this relationship, on the left, we impose the condition that there is no interaction effect and then plot the resulting π_{HS}/π_{K12} against the poverty rate where the present value is evaluated at the average K12 spending levels, 10% above this average and 10% below this average.

As one would expect, on the left panel, this ratio is falling with the poverty rate. This reflects the fact that K12 spending has a larger effect on poor children so that the average benefits of K12 spending are larger in higher poverty areas. Also as expected, (even where there is dynamic complementarity) the ratio is higher when evaluated at higher levels of K12 spending. Interestingly, with no dynamic complementarity, the relative marginal benefit of rolling out a Head Start center lies below that of K12 spending so long as the poverty rate is above about 20 percent. With no dynamic complementarity, this is true even in areas that spend 10% above average in the K12 system. To illustrate how dynamic complementarity affects these ratios, we allow for dynamic complementary (i.e., using the estimated interaction term from Table 2) and then evaluate these same ratios (right panel). Evaluated at the average, the basic pattern is the same. However, with dynamic complementarity, the ratios are very different at K12 spending levels 10% above and below the average. Where complementarities exist, in areas that spending 10% higher than average in the K12 system, this ratio lies above 1 at all poverty levels, so that the marginal impact of rolling out Head Start on average wages is larger than the effects of spending that same money in the K12 system. The flipside of this result is that in areas that spend less than 10% lower than the average, this ratio lies below 1 for all poverty levels. This means that in areas with low levels of K12 spending, the marginal dollar is better spent in the K12 system than on Head Start.

In essence, these patterns support the idea that, when such dynamic complementarities exist between early and late human capital investments, in some locations, 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 $4,500 per-pupil (about 10% above the average K12 spending level), the marginal dollar spent on Head Start led to between 1.5 and 2.5 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 8 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 and Conclusions

This study provides new evidence on the life-cycle effects of Head Start and K12 school spending. We explore dynamic complementarities between human capital investments made in pre-school and those that subsequently occur in the K12 system. We use children's differential exposure to Head Start spending (at age 4) and court-ordered school finance reforms (SFRs) (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 poor 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 from K12 school districts that spent at the 75th percentile of the distribution than those from K12 school districts that spent at the 25th percentile. Similarly, the benefits of K12 school-spending increases on adult outcomes were larger among poor children who were exposed to higher levels of 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 independent effects of the two investments 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.[39] 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 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. At the same time, the returns to education have increased, so the consequences of access to high-quality human capital investments from preK-12 are large.

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 explain 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

[39] 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.

which they are implemented. The findings point to the critical role early-life investments can play in narrowing long-run gaps in well-being, and they also highlight the importance of sustained investments in the skills of disadvantaged youth.

References

Aizer Anna, 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.

Neidell, Matthew and Jane Waldfogel. “Cognitive and Non-Cognitive Peer Effects in Early Education,” *Review of Economics and Statistics*, 92(3), 2010.

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.

Office of Child Development. 1970, “Project Head Start 1968; the Development of a Program” Office of Child Development (DREW), Washington, D.C.

Office of Economic Opportunity. 1967. Project Head Start. *The Quiet Revolution: Second Annual Report of the Office of Economic Opportunity*. Washington, D.C.: Government Printing Office.

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, 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=15,232)	Poor Child (N=6,373)	Non-Poor Child (N=8,859)
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 (average, ages 5-17)	$4,366	$4,031	$4,470
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 (average, ages 12-17)	$54,488	$22,520	$65,130
Income-to-needs ratio (average, 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 kid" is defined here as children whose parents were in the bottom quartile of the income distribution (approximately 80% of whom were below the poverty line). Analysis sample includes 15,232 individuals (218,594 person-year observations ages 20-50), from 4,990 childhood families, 1,427 school districts, 1,120 childhood counties and all 50 states.
*Child-specific pre-K attendance & Head Start program participation info collected retrospectively in 1995 survey IW.

Table 2: *Marginal Effects of Head Start Spending and Public Per-Pupil Spending and Their Interaction: Poor Children*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Prob(High School Grad)		Years of Completed Education		Ln(Wage), ages 20-50		Annual Incidence of Poverty, age 20-50		Prob(Ever Incarcerated)	
	DiD-2SLS	2SLS-IV	DiD-2SLS	2SLS-IV	DiD-2SLS	2SLS-IV	DiD-2SLS	2SLS-IV	DiD-2SLS	2SLS-IV
Head Start Spending$_{(age\ 4)}$	0.02503	0.04089	0.07721	0.2255	0.02334	0.03615	-0.01808	-0.02576	-0.006002	-0.02024
	(0.006942)	(0.02453)	(0.01992)	(0.1212)	(0.004503)	(0.01956)	(0.005302)	(0.01385)	(0.003494)	(0.01082)
(SFR) Instrumented Ln(PPE)$_{(age\ 5\text{-}17)}$	1.1016	1.4163	4.0399	4.0218	2.0561	1.2596	-0.7923	-0.7971	-0.8080	-1.1822
	(0.3268)	(0.3390)	(1.6751)	(1.7856)	(0.4348)	(0.2690)	(0.2969)	(0.2903)	(0.3397)	(0.4550)
Head Start Spending$_{(age\ 4)}$*ln(PPE)$_{(age\ 5\text{-}17)}$	0.1012	0.2273	0.6460	0.8345	0.1698	0.2561	-0.1079	-0.1852	-0.05169	-0.1808
	(0.05454)	(0.06518)	(0.2354)	(0.4824)	(0.06985)	(0.07191)	(0.04267)	(0.05038)	(0.02777)	(0.1076)
Marginal Effects of 10% increase in K12 Spending by Head Start access:										
No Head Start$_{(age\ 4)}$	0.0673	0.0455	0.1307	0.0492	0.1338	0.0176	-0.0336	-0.0014	-0.0589	-0.0418
	(0.0236)	(0.0316)	(0.1274)	(0.1064)	(0.0349)	(0.0219)	(0.0301)	(0.0153)	(0.0283)	(0.0193)
Head Start Center access$_{(age\ 4)}$	0.1102	0.1416	0.4040	0.4022	0.2056	0.1260	-0.0792	-0.0797	-0.0808	-0.1182
	(0.0327)	(0.0339)	(0.1675)	(0.1786)	(0.0435)	(0.0269)	(0.0297)	(0.0290)	(0.0340)	(0.0455)
Marginal Effects of Head Start with 10% increase or decrease in K12 Spending:										
w/10% decrease	0.0630	0.0768	0.0533	0.6010	0.0269	0.0446	-0.0308	-0.0306	-0.0035	-0.0092
	(0.0481)	(0.1169)	(0.1393)	(0.5937)	(0.0284)	(0.0921)	(0.0206)	(0.0568)	(0.0229)	(0.0487)
Average	0.1059	0.1730	0.3266	0.9540	0.0987	0.1529	-0.0765	-0.1090	-0.0254	-0.0856
	(0.0294)	(0.1038)	(0.0843)	(0.5129)	(0.0190)	(0.08275)	(0.0224)	(0.0586)	(0.0148)	(0.0457)
w/10% increase	0.1487	0.2691	0.5999	1.3070	0.1706	0.2613	-0.1221	-0.1873	-0.0473	-0.1621
	(0.0217)	(0.0968)	(0.1209)	(0.5068)	(0.0408)	(0.0841)	(0.0351)	(0.0674)	(0.0138)	(0.0772)
Number of Person-year Observations	--	--	--	--	55,706	55,706	88,124	88,124	--	--
Number of Children	5,419	5,419	5,419	5,419	5,613	5,613	6,373	6,373	4,536	4,536

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

Data: PSID geocode Data (1968-2015), matched with childhood school and neighborhood characteristics. Sample includes all individuals born 1950-1976 whose parents were in the bottom quartile of the income distribution, and who have been followed into adulthood.

Models: (non-Instrumented & Instrumented) Head Start Spending per poor 4-year old at age 4 in the county and instrumented ln(school district per-pupil spending during ages 5-17) are centered around their respective means, to facilitate interpretation of the main effects evaluated at roughly the mean; the average SFR-induced increase in school-age spending is about 10%. Results are based on 2SLS-IV models that 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; controls for 1960 county characteristics (poverty rate,% black, education,% urban, population size, each interacted with linear cohort trends; controls for county-level per-capita gov't safety net expenditures average during childhood; and controls for childhood family characteristics (parental income/education, mother's marital status at birth, birth weight, gender), and age (cubic). The first-stage model includes 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 instrument used for Head Start spending per poor 4-year old is an indicator for whether there was any Head Start center in the county at age 4 (based on the program's rollout timing variation only). There exists a significant first-stage. The marginal effects related to Head Start access are based on the average county Head Start spending when there is a center (~$4,230 (in real 2000 dollars))--i.e., marginal effects are evaluated for roughly a $4K increase in Head Start spending (to contrast impact of having access vs no access to Head Start center).

Table 3: *Marginal Effects of Head Start Spending and Public Per-Pupil Spending and Their Interaction: Non Poor Children*

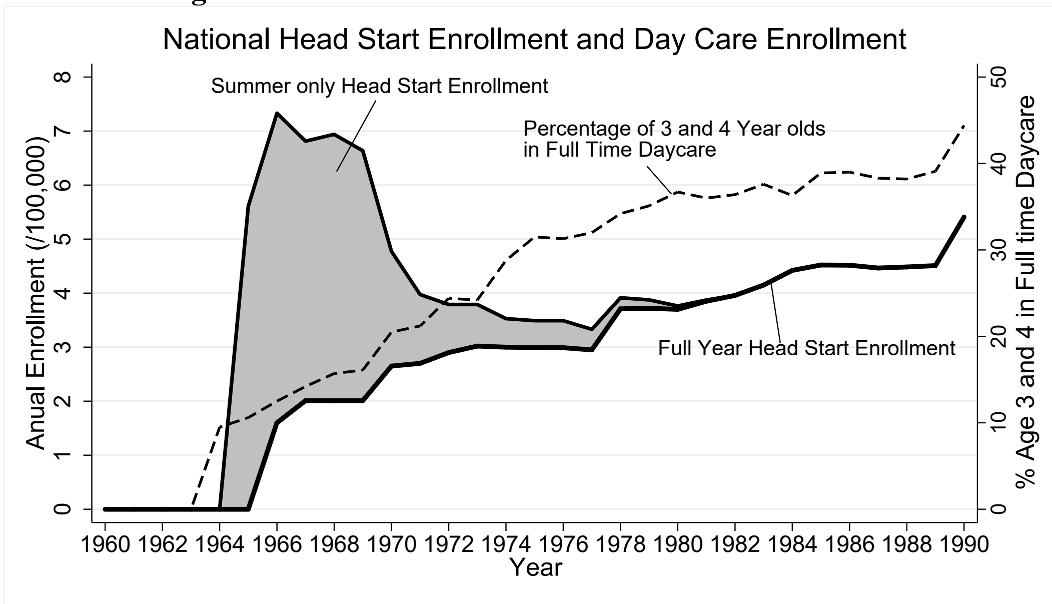
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Prob(High School Grad)		Years of Completed Education		Ln(Wage), ages 20-50		Annual Incidence of Poverty, age 20-50		Prob(Ever Incarcerated)	
	DiD-2SLS	2SLS-IV	DiD-2SLS	2SLS-IV	DiD-2SLS	2SLS-IV	DiD-2SLS	2SLS-IV	DiD-2SLS	2SLS-IV
Head Start Spending$_{(age\ 4)}$	0.000014	-0.02227	0.008866	-0.05426	0.006901	0.01932	-0.000085	-0.008452	-0.001274	0.000937
	(0.003432)	(0.01864)	(0.01635)	(0.1071)	(0.005408)	(0.02433)	(0.001716)	(0.005692)	(0.001705)	(0.006974)
(SFR) Instrumented Ln(PPE)$_{(age\ 5\text{-}17)}$	0.2386	0.4671	2.4192	2.1565	0.7351	0.4155	-0.1383	-0.1868	-0.09837	0.1298
	(0.1197)	(0.2351)	(1.1645)	(1.5314)	(0.3035)	(0.2366)	(0.06316)	(0.1304)	(0.2161)	(0.2758)
Head Start Spending$_{(age\ 4)}$*ln(PPE)$_{(age}$	0.01688	0.09666	0.02972	0.4144	0.02577	-0.01603	0.005707	-0.000716	-0.02568	-0.01128
	(0.02347)	(0.08062)	(0.1937)	(0.3706)	(0.03090)	(0.1020)	(0.01459)	(0.04094)	(0.02271)	(0.06604)
Number of Person-year Observations	--	--	--	--	90,771	90,771	130,470	130,470	--	--
Number of Children	7,983	7,983	7,983	7,983	8,195	8,195	8,859	8,859	5,140	5,140

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

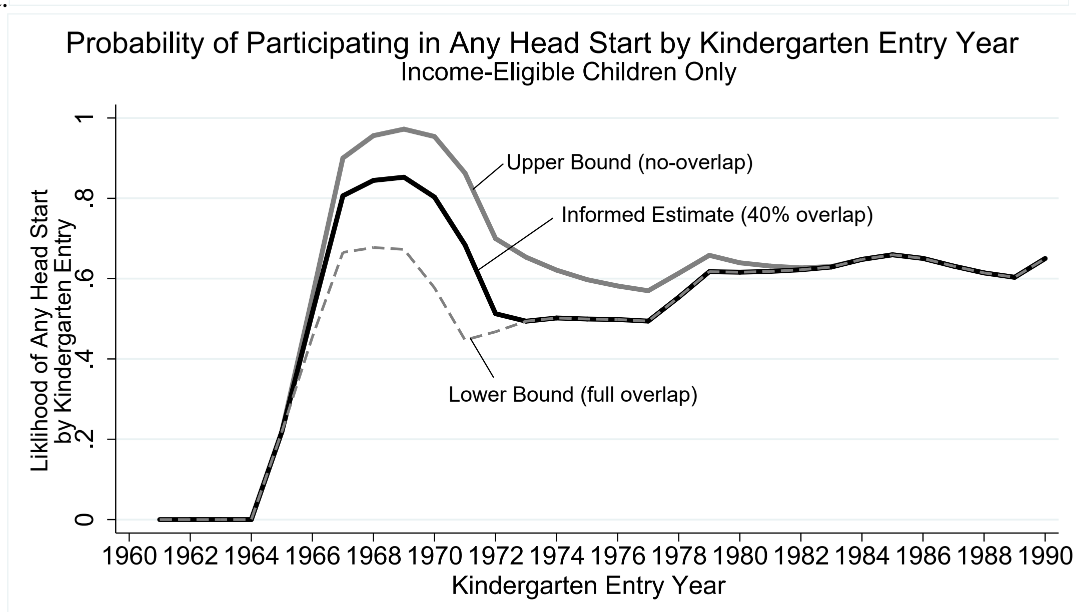
Data: PSID geocode Data (1968-2015), matched with childhood school and neighborhood characteristics. Sample includes all individuals born 1950-1976 whose parents were NOT in the bottom quartile of the income distribution, and who have been followed into adulthood.

Models: (non-Instrumented & Instrumented) Head Start Spending per poor 4-year old at age 4 in the county and instrumented ln(school district per-pupil spending during ages 5-17) are centered around their respective means, to facilitate interpretation of the main effects evaluated at roughly the mean; the average SFR-induced increase in school-age spending is about 10%. Results are based on 2SLS-IV models that 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; controls for 1960 county characteristics (poverty rate,% black, education,% urban, population size, each interacted with linear cohort trends; controls for county-level per-capita gov't safety net expenditures average during childhood; and controls for childhood family characteristics (parental income/education, mother's marital status at birth, birth weight, gender), and age (cubic). The first-stage model includes 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 instrument used for Head Start spending per poor 4-year old is an indicator for whether there was any Head Start center in the county at age 4 (based on the program's rollout timing variation only). There exists a significant first-stage. The marginal effects related to Head Start access are based on the average county Head Start spending when there is a center (~$4,230 (in real 2000 dollars))--i.e., marginal effects are evaluated for roughly a $4K increase in Head Start spending (to contrast impact of having access vs no access to Head Start center).

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



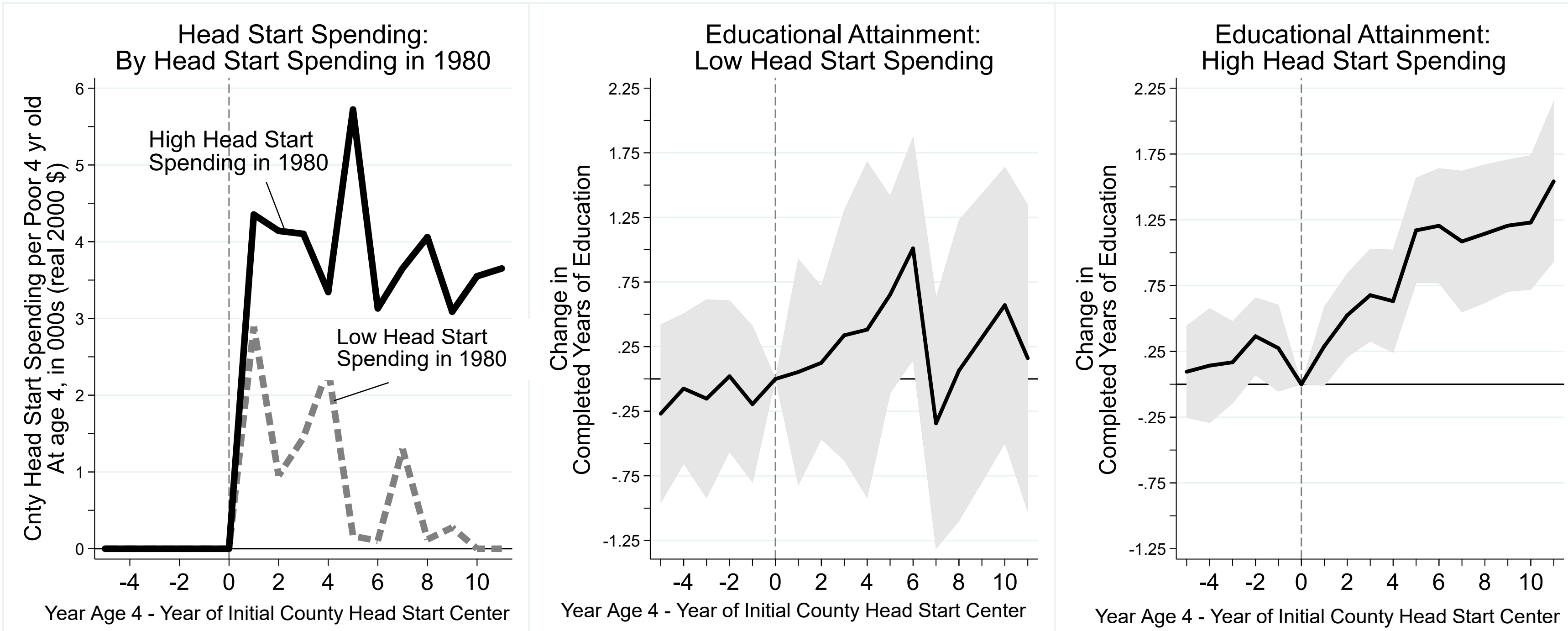
a.



b.

Notes: National counts of 3,4, and 5-year-olds are derived from Integrated Public Use Microdata Series (IPUMS) decennial censuses from 1960 to 1990. Counts for non-census years are completed using linear interpolation. The percentage of children ages 3 and 4 who are enrolled in full-year daycare are as reported in the Current Population Survey (link: http://www.census.gov/hhes/school/data/cps/historical/.) 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). In panel b, the participation rate is the cumulative probability of enrolling in Head Start across all age-eligible years prior to Kindergarten entry (note: this is not the same as the fraction of eligible enrollees in a given year, but is the sum of these annual probabilities across age-eligible years prior to kindergarten entry. The upper bound assumes that each enrollee in full years and summer only programs is unique (no overlap). The lower bound assumes that all full-year enrollees, were possible, were also in a summer program (full overlap). The informed estimates assume that, where possible, 40 percent of the full year enrollees were previously summer enrollees.

Figure 2: *Evolution on Head Start Spending and Educational Attainment at Rollout (Poor Children)*

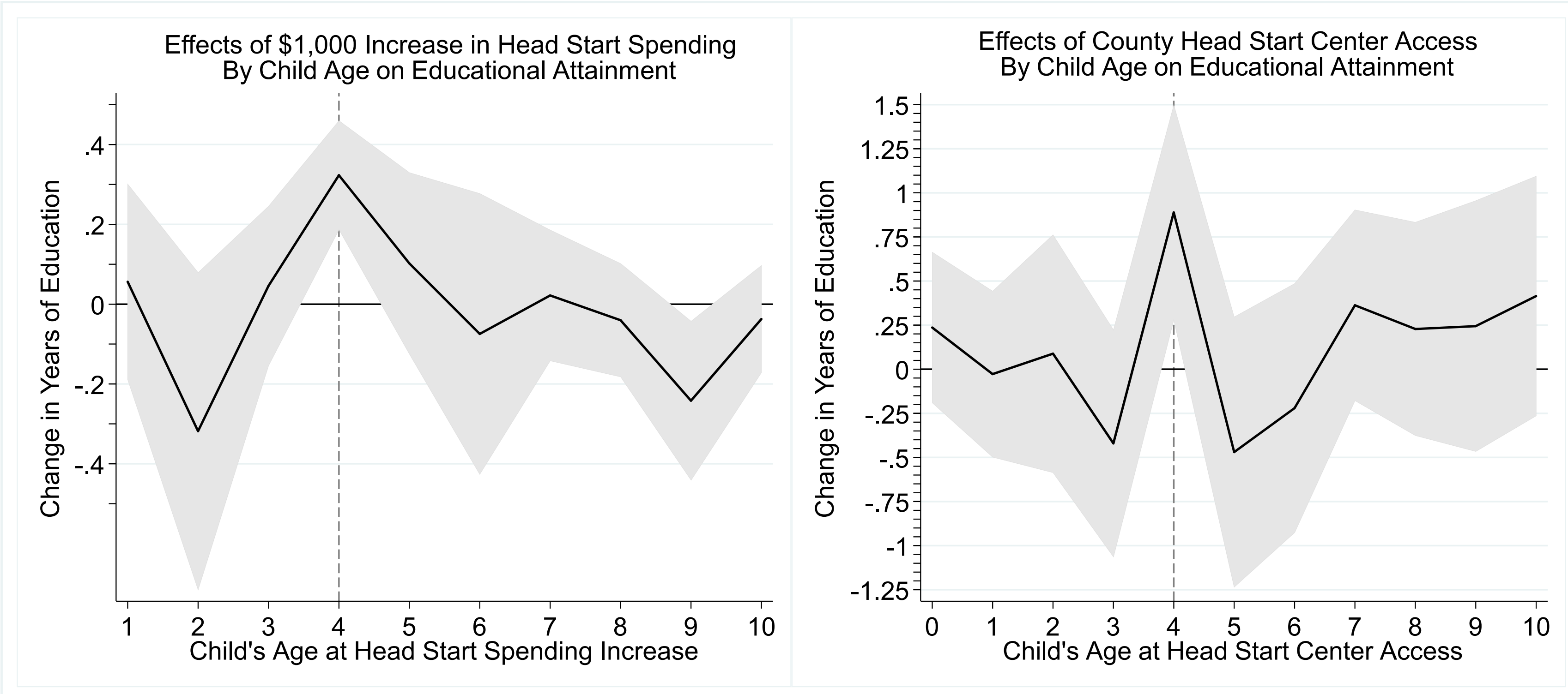


Data: Analysis sample includes PSID individuals born 1950-1976 who have been followed into adulthood. "High Head Start spending" is defined here as counties in the top quartile of Head Start spending among all US counties after rollout; "Low Head Start spending" defined here as bottom quartile of Head 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 figures present the event-study plots for both high and low spending counties (in 1980). The shaded grey region in the event study plots for years of education depict the 90% confidence interval for each event-year. The models include childhood county fixed effects, race*census division-specific birth year trends; controls for 1960 county characteristics (poverty rate,% black, education,% 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).

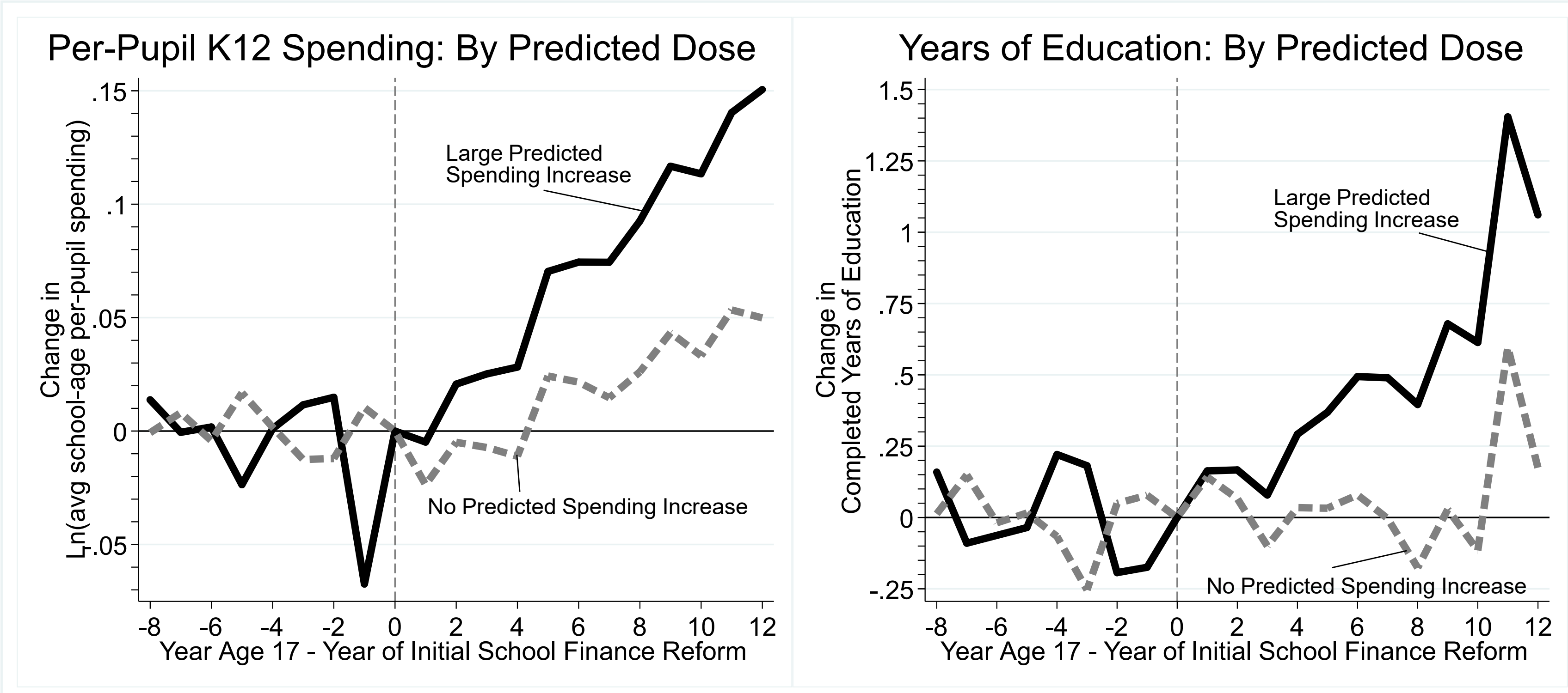
Figure 3:

Effects of Head Start Spending and Access: by Age of Spending and Access (Poor Children)



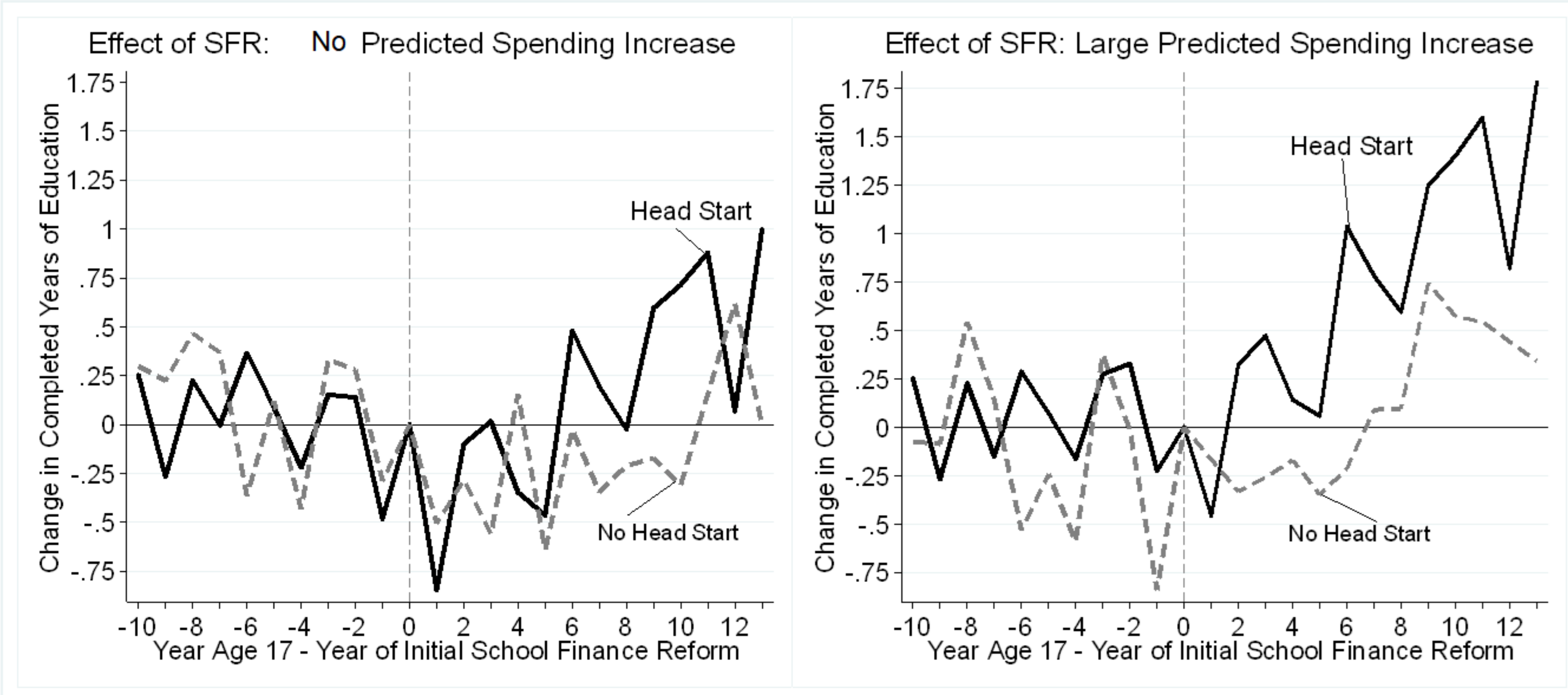
These figures present the marginal effects of Head Start spending in an individual's childhood county at different ages, conditional on the level of Head Start spending in the childhood county at age 4 (when such spending should have an effect). The shaded grey region in the event study plots depict the 90% confidence interval for each rollout age estimate. 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 a model with no other ages included. For the regression estimates underlying this model for years of education attained, see Appendix Table H6.

Figure 4: *Evolution of K12 Spending and Educational Attainment after SFR Reform (All Children)*



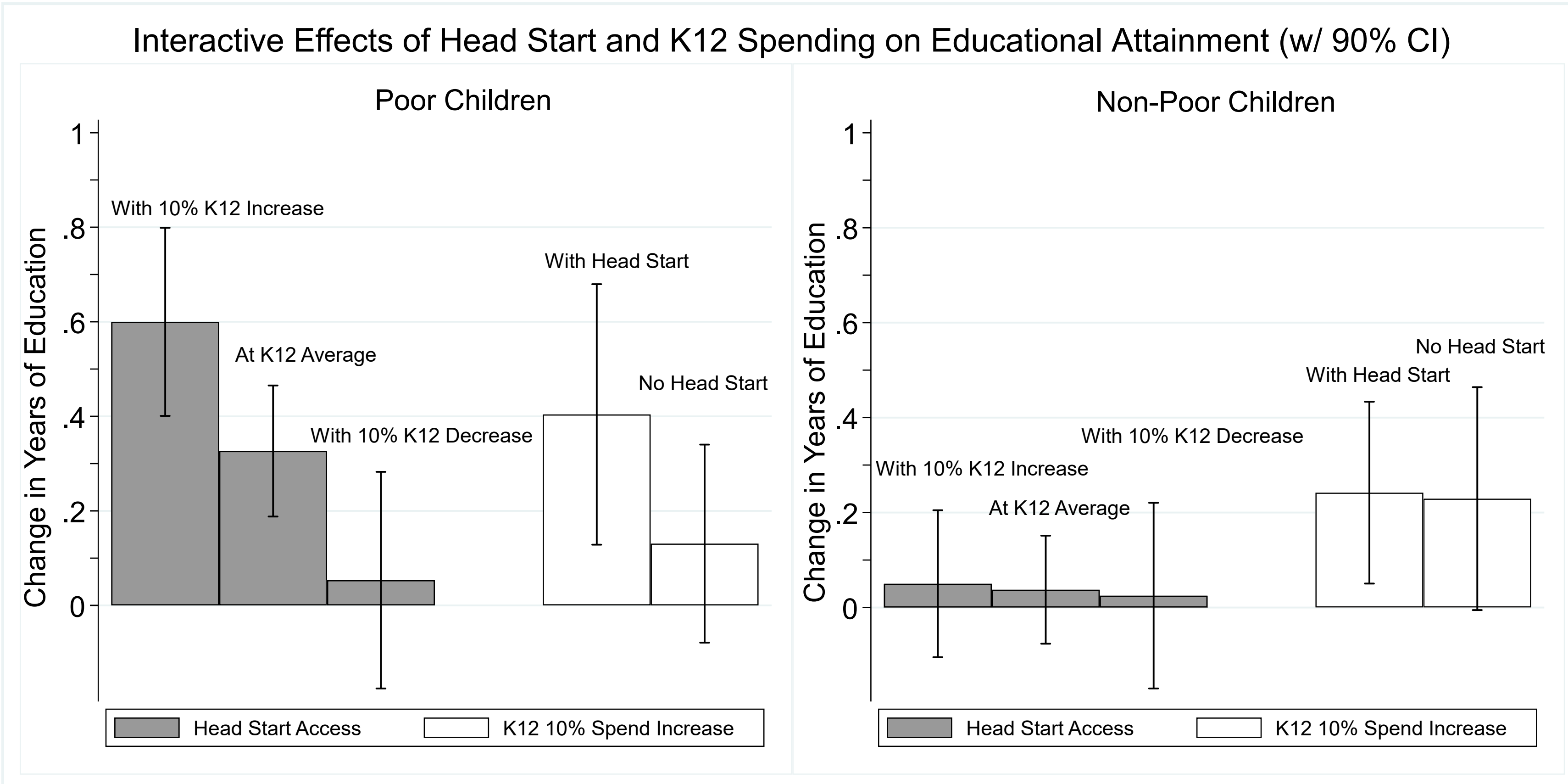
Models: The event study figures use school district's predicted reform-induced change in spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories—the solid black line shows estimated effects for districts with 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,% black, education,% 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 5: *Effect of K12 Spending on Year of Completed Education: by Head Start Exposure Status (Poor Children)*



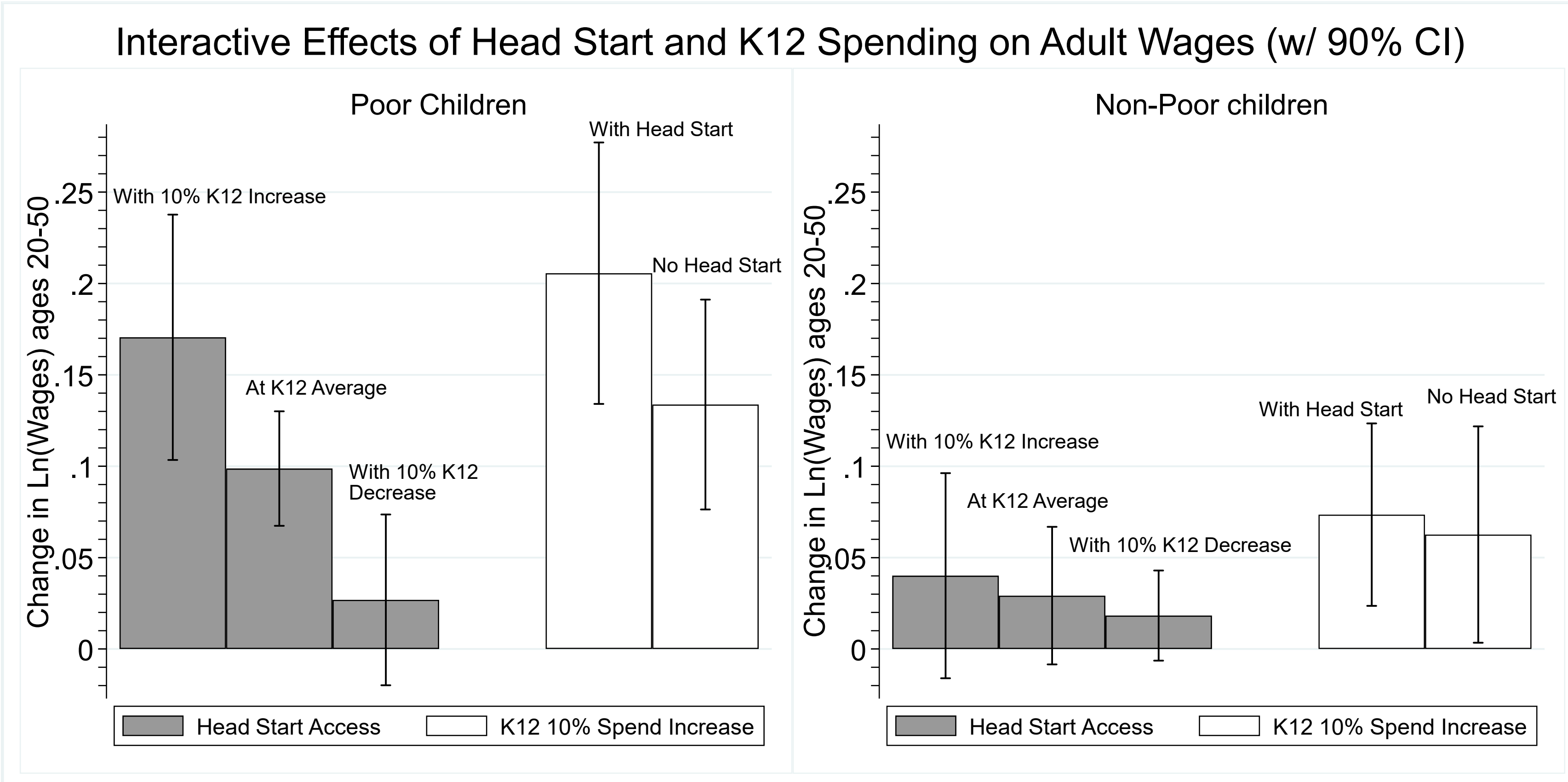
Models: The event study figures use school district's predicted reform-induced change in spending based on the timing and type of court-ordered reform interacted with 1970 (within-state) district income and spending percentile categories--right panel shows estimated effects for districts with a predicted 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,% black, education,% 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 Head Start Spending by K12 spending Levels and vice versa on Educational Attainment*



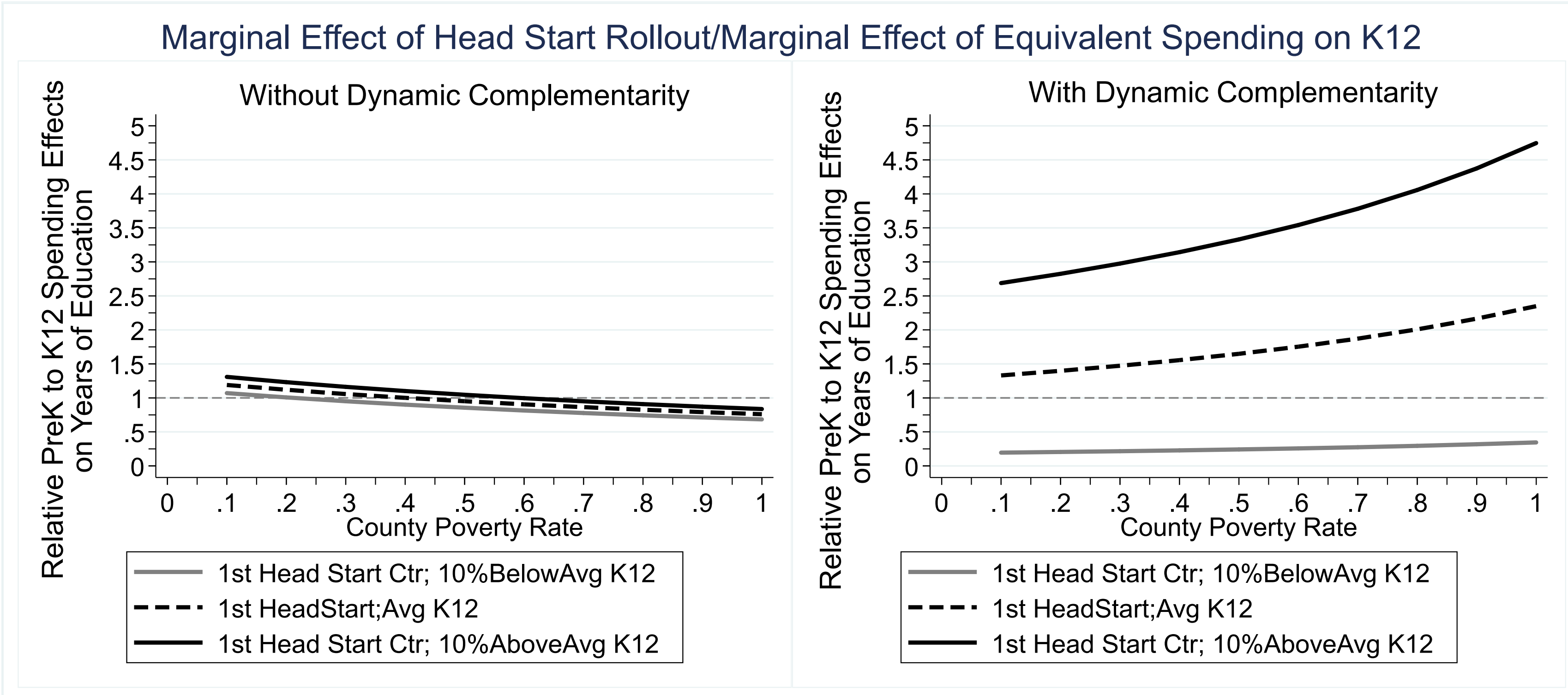
Note: The reported marginal Effects based upon 2SLS-Difference-in-Difference model results are presented in columns 3 from Tables 2 and 3. The reported marginal effects and the standard errors were computed using the delta method.

Figure 7: *Effect of Head Start Spending by K12 spending Levels and vice versa on Wages*



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.

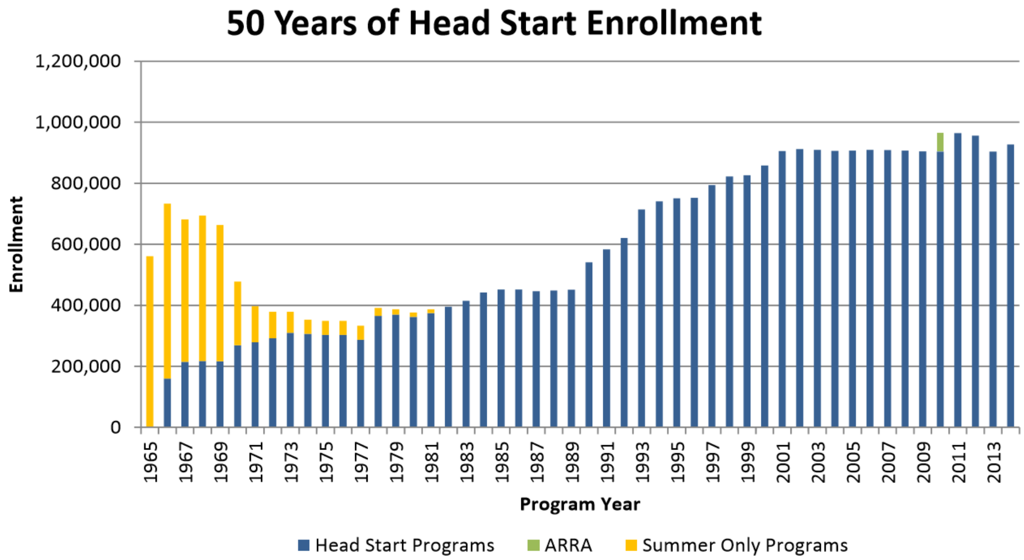
Figure 8: *Ratio between the Effect of Head Start Spending and K12 spending Levels by Poverty Level in the County*



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.

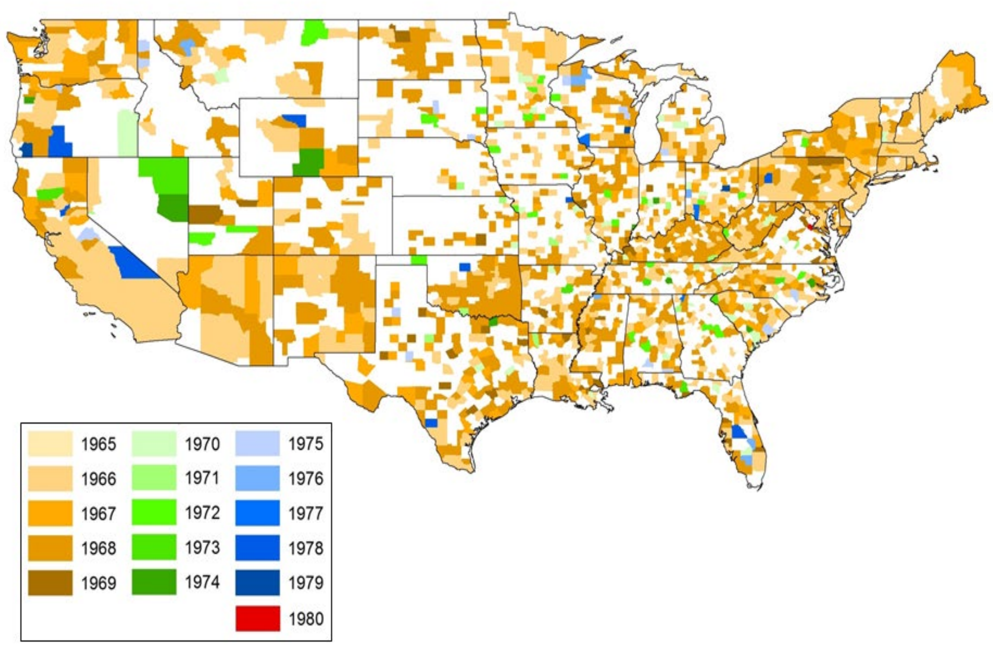
Appendix A
Additional Tables

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

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 poor 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% increase in spending per poor four-year-old is associated with only a 0.243%.

Table B1:

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-	
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

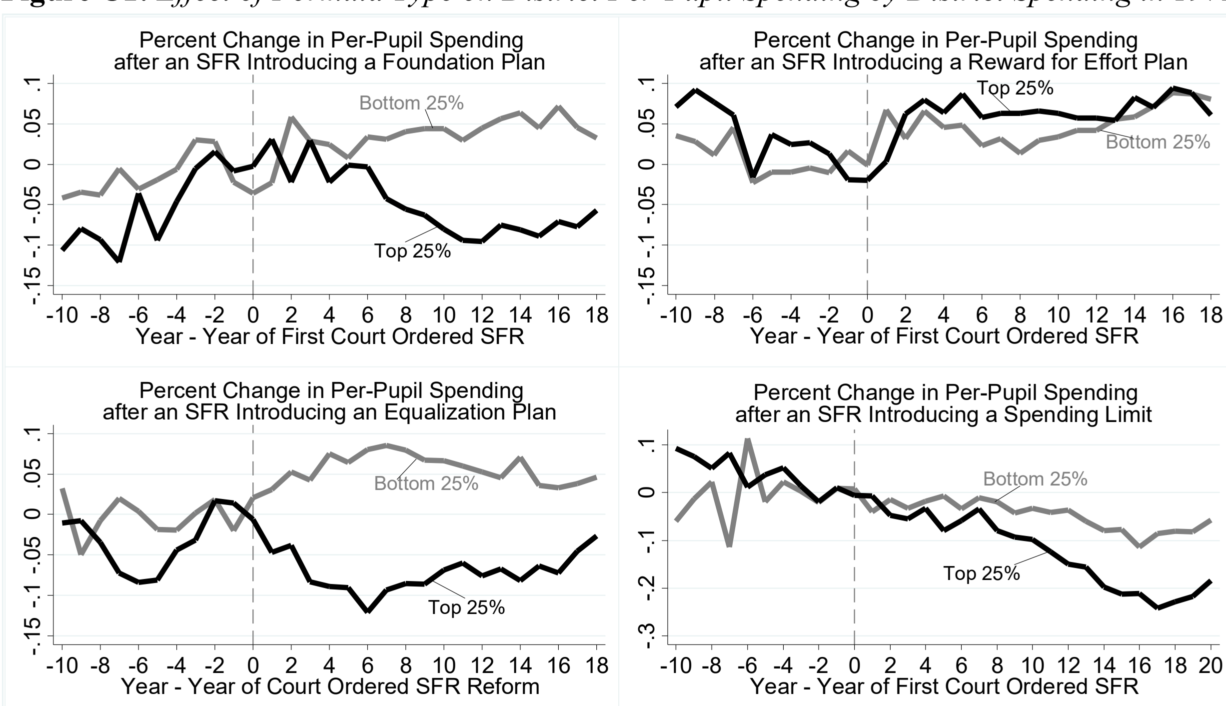
** p<0.01, * p<0.05, + p<0.1

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

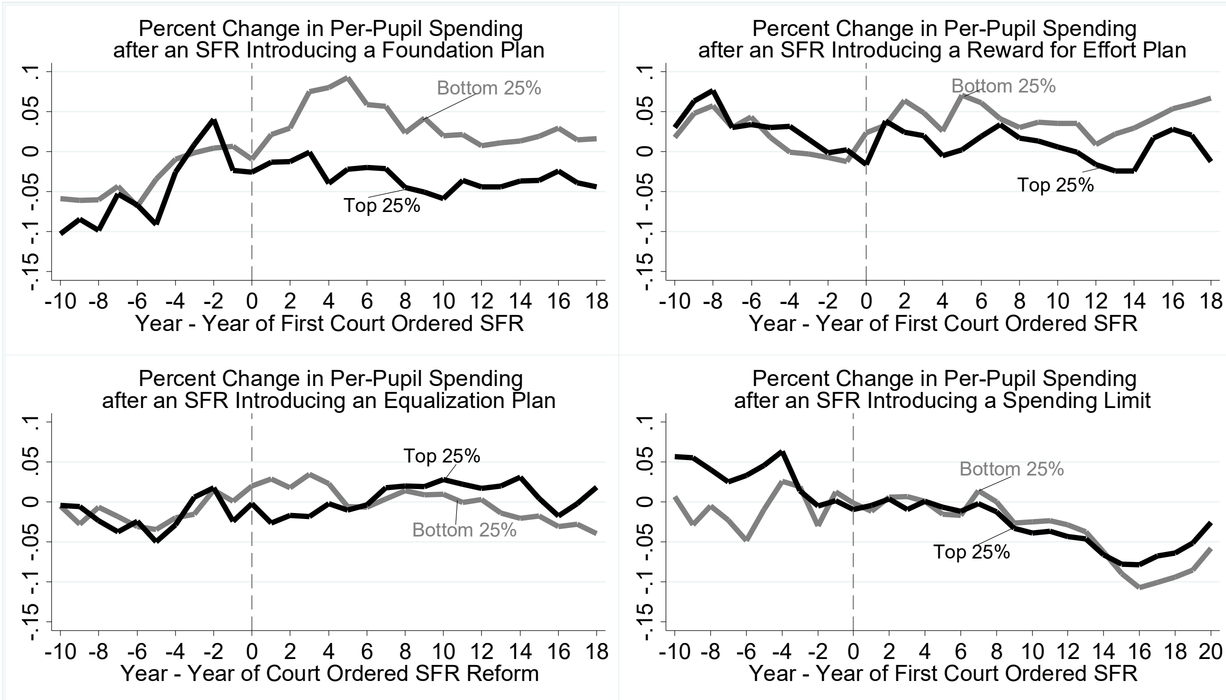
To illustrate how the introduction of different formula types affected districts by pre-reform income and spending levels, we replicate the analysis in Jackson Johnson and Persico (2016). Figures C1 and C2 present event-study plots of the natural log of per-pupil spending at the district level (after 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 C1 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 C2 presents similar plots for districts in the top and bottom quartiles of the state income distribution in 1963. Figures C1 and C2 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.

Figure C1: *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% 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 C2: *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% 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.

Appendix D: *Panel Study of Income Dynamics (PSID, 1968-2015)*

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.

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 D1). In particular, among original sample children, baseline 1968 family and county characteristics do not jointly significantly predict the likelihood of attrition or the likelihood of being observed as an adult.

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).

Matching PSID Individuals to their Childhood School Districts

We use the confidential restricted-use geocode PSID data that includes census block identifiers that correspond with childhood respondent addresses. We match respondent earliest childhood residential location (typically, 1968) to school districts via the combination of GIS mapping methods and school-to-census tract relationship files. 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. 1970 street addresses for schools are obtained from the Elementary and Secondary General Information System (ELSEGIS). Using GIS software, we locate these schools using the 2000 census electronic road maps (http://www.esri.com/data/download/census2000_tigerline/). We use a crosswalk of census tract identifiers across 1970/1980/1990/2000/2010 censuses (since the definitions of neighborhoods change over time), and 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.

Table D1: *PSID Analysis Tests of Sample Attrition***.**

	Dependent variable:		
	Probability(original sample child observed in adulthood)		
	All	**Poor Children**	**Non-Poor Children**
	(1)	(2)	(3)
1968 Family & County Characteristics:			
Black (ref cat: white)	0.000241	-0.004006	0.003434
	(0.01807)	(0.02313)	(0.03497)
Family income-to-needs ratio	0.001694	0.007748	-0.0001174
	(0.006366)	(0.02573)	(0.008722)
Female-headed household	-0.03048	-0.01692	-0.06152
	(0.01874)	(0.02210)	(0.04078)
Number of children	0.006430	0.007830	0.008396
	(0.004636)	(0.006050)	(0.008612)
Parental education (ref cat: high school grad):			
High school dropout	-0.01927	-0.01291	-0.02987
	(0.01659)	(0.02264)	(0.02559)
Attended college	0.01494	-0.03191	0.03238
	(0.02214)	(0.04690)	(0.02446)
Household annual food expenditures	-0.0002157	-0.0002115	-0.0003181
	(0.0001734)	(0.0003127)	(0.0002228)
Parental expectations for achievement, index	-0.002894	-0.001608	-0.003156
	(0.003762)	(0.005337)	(0.005312)
County unemployment rate	0.002809	0.01084	-0.008268
	(0.008466)	(0.01264)	(0.01120)
County public assistance expenditures per capita	0.0007209	0.001196	0.0004415
	(0.002912)	(0.004356)	(0.004037)
Region (ref cat: South):			
Northeast	0.0002421	-0.02104	0.03062
	(0.01943)	(0.02879)	(0.02690)
Midwest	0.0004922	-0.008726	0.01732
	(0.01833)	(0.02709)	(0.02628)
West	-0.004869	-0.03543	0.02812
	(0.02173)	(0.03525)	(0.02809)
F-test of joint significance (p-value)	F-stat=0.83;	F-stat=0.42;	F-stat=1.12;
Proportion of original sample children observed in adulthood	0.766	0.753	0.766

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Data: PSID geocode Data (1968-2015): Analysis sample includes PSID original sample children born 1950-76. 76.6% of these children have been followed into adulthood, and are included in analysis sample in main results presented in Tables 2 and 3. This Table shows no evidence of selective attrition based on 1968 childhood family and county characteristics. (p-value of F-test of joint significance of vars = 0.8301).

Appendix E

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 dataset 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.[40] Note that the SEERS data are not broken down by poverty and age. As such, we obtain the 1970 county-level child poverty rate via the 1970 Census (ICPSR) data and multiply this by the county-level number of 4-year olds, which together provides an accurate estimate of the number of poor 4-year olds in each county (assuming county-level child poverty rates do not differ greatly by child age).

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

[40] 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.

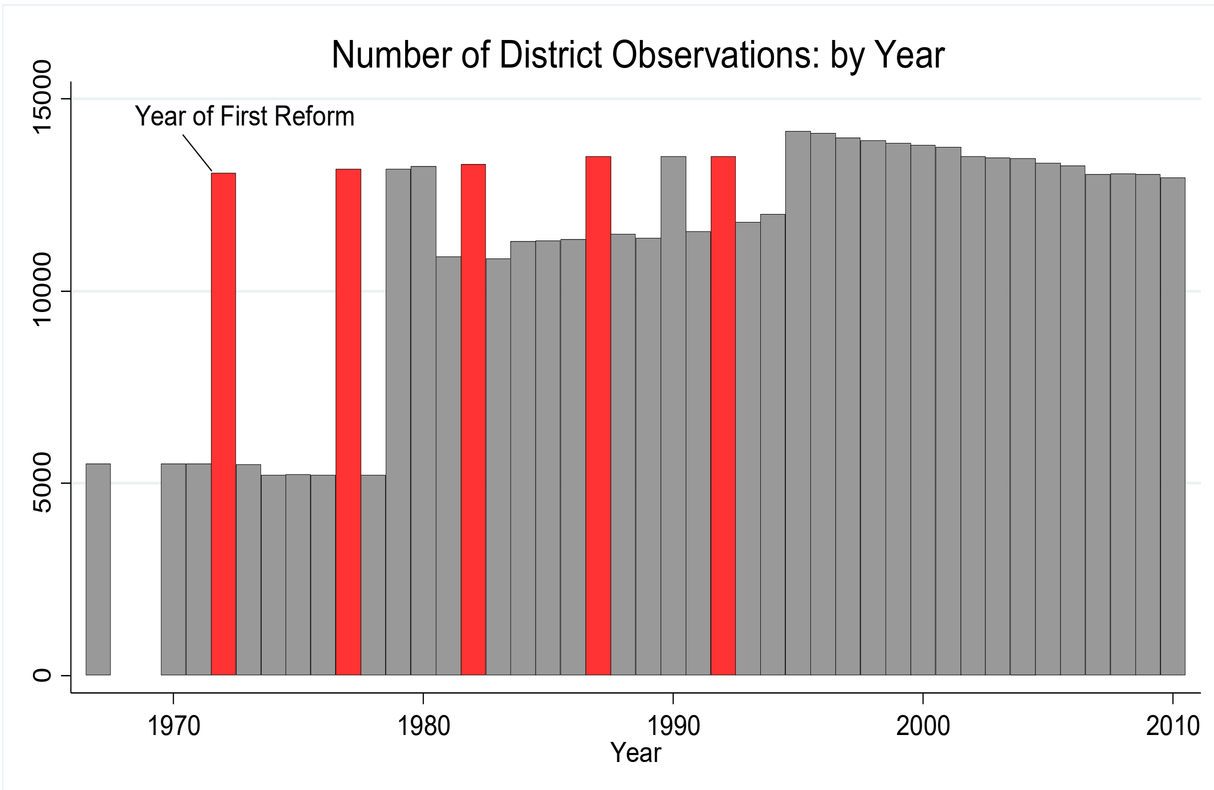
example, this database includes corrected statistical weights that have been standardized across years, which had not been done previously. Furthermore, although most governments retain the ID number they are assigned originally, there are circumstances that result in a government's ID being changed. Since a major purpose of the INDFIN database is tracking government finances 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 E1 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 than that used previously, it is not perfect. As shown in Appendix Figure E1, for years, 1967, 1970, 1971, 1973, 1974, 1975, 1976, and 1978 only about 40% of districts are present (often larger districts). After 1979 almost all districts are included.

Figure E1: *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 United 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).

Data on Other Policies and Additional Controls

The data we use for other controls 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 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).

Appendix F

Predicting Dosage

The prediction of $dose_d$ is obtained in two steps. We discuss each step in turn below.

Step 1

First, using district-by-birth-cohort data for the full universe of districts (not only those represented in the PSID), we use flexible Difference in Difference regression models to predict how school spending in each district responded to the passage of a court-ordered reform based on (a) the type of reform introduced after the court order interacted with the district's school spending levels prior to reforms, and (b) the type of reform introduced after the court order interacted with the district's income level prior to reforms.

To do this, we estimate [F1] where all common variables are defined as in [4] and [5]. In [F1], *T* is event time and is the year an individual turned 17 minus the years of the first court-ordered SFR in their state of birth. Accordingly, *T* is 0 for those who turned 17 the years of a SFR and are essentially not exposed, it is -2 for those who turned 19 during the year of an SFR so that they graduated high school 2 years before the SFR, and T would be 5 for individuals who turned 17 f years after the first SFR in their state of birth. This exposure measure varies at the state birth-cohort level and goes from -20 (those who were age 17 twenty years before 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). In [F1], $I_{F,d}$ 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.

[F1] $$\ln(PPE_{5-17})_{idb} = \sum_{Q_{ppe=1}}^{4} \sum_{T=-20}^{20} \left(I_{T_{idb}=T} \times I_{Q_{ppe72,d}=Q_{ppe}} \right) \cdot \alpha_{T,Q_{ppe}} + \sum_{F=1}^{5} \sum_{Q_{ppe=1}}^{4} \sum_{T=-20}^{20} \left(I_{T_{idb}=T} \times I_{Q_{inc69,d}=Q_{ppe}} \times I_{F,d} \right) \cdot \alpha_{T,Q_{inc},F} + \Pi C_{idb} + \theta_{d3} + \theta_{b3} + \varphi_{idb}.$$

The variables $\sum_{Q_{ppe=1}}^{4} \sum_{T=-20}^{20} \left(I_{T_{idb}=T} \times I_{Q_{ppe72,d}=Q_{ppe}} \right)$ are the set of interactions between the quartile of district *d* in the state distribution of per-pupil spending in 1972, and exposure to an SFR. Accordingly, 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, $\sum_{F=1}^{5} \sum_{Q_{ppe=1}}^{4} \sum_{T=-20}^{20} \left(I_{T_{idb}=T} \times I_{Q_{ppe72,d}=Q_{ppe}} \times I_{F,d} \right)$ are the set of interactions between the type of reform, the quartile of district *d* in the state distribution of median income in 1969, and exposure to an SFR. Accordingly, 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.

Step 2

In the second step, we take the estimates from estimation of [F1] to summarize how a given districts per pupil spending is likely to change after the introduction of a courted ordered SFR in their state. That is, 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 (i.e., those six cohorts exposed to an SFR for between 3 and 8 years). To assuage any concerns that this age range choice is arbitrary, note that our results are similar when using other ages such as ages 10 to 17 or 5 to 17.[41] Formally, our predicted district-specific dose effect based on [F1] is

[F2] $\widehat{\text{dose}}_d = \Big[\sum_{Q_{ppe}=1}^{4}\sum_{T=3}^{8}\left(I_{T_{idb}=T} \times I_{Q_{ppe72,d}=Q_{ppe}}\right)\cdot \hat{\alpha}_{T,Q_{ppe}} + \sum_{F=1}^{5}\sum_{Q_{ppe}=1}^{4}\sum_{T=3}^{8}\left(I_{T_{idb}=T} \times I_{Q_{ppe72,d}=Q_{ppe}} \times I_{F,d}\right)\cdot \hat{\alpha}_{T,Q_{inc},F}\Big]/6$

By using the predicted values, $\widehat{\text{dose}}_d$, from [F2] 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.[42] 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.

NOTE: We estimate our main models excluding $\widehat{\text{dose}}_d$, (i.e. using only variation due to exposure to an SFR) and the results are very similar. See appendix Table H2.

[41] We chose this age range because it included enough years (i.e., 5) to not be sensitive to random fluctuations in the high frequency data, and because it occurred relatively soon after the passage of a court-ordered reform (these exclude the first two years following a court order as there was typically a two-year delay in legislative implementation of SFRs following a court order, with limited spending changes in the year immediately after.

[42] 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 G

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 important ways. First, the data are retrospective data collected in the 1995 survey wave based on questions that asked adults about their early childhood experiences and whether they had ever participated in a Head Start program. 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. Potential recall bias may be particularly problematic for Head Start participation during the ramp-up period during which most of our variation is derived. This is due to 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 were in the summer-only programs) between 1965 and 1970 are not reliably recorded in the participation survey responses reported in the PSID. As a result, relating increases in Head Start spending to retrospectively reported Head Start participation in the PSID might drastically understate the effects of Head Start spending on enrollment in the program.

Having discussed 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. 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 18.02 percentage points. This implies that for the average county that spends $4,320 per poor four-year-old, Head Start participation is estimated to increase by 18.2*4.32=78.6 percentage points. We also estimate the effect of Head Start rollout (e.i. if a Head Start center was available in the individual's county of birth when they were 4 years old) on participation. The marginal effect from the conditional logit model is 0.797. This is almost identical to the implied effect from the spending specification and suggests that rollout increased Head Start participation by roughly 80 percentage points among children from poor families. Because the conditional logit model requires that there be some variation in the outcome within each county, we cannot run the conditional logit models on the non-poor population because the vast majority of counties do not have any non-poor children in Head Start.

This implied participation effect at rollout of about 80 percentage points for poor children is very similar to our assumed participation effect of 75% for any Head Start program, somewhat larger than our assumed participation effect of about 63% for full-time Head Start, and roughly 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 G1:
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 Head Start)
	Conditional Marginal Effects, evaluated at means	
County Head Start Spending per Poor 4-year old$_{\text{(age 4)}}$ (in 000s)	0.1802***	
	(0.0205)	
Head Start Center(age 4)		0.7972***
		(0.0500)
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 H

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].

[H1] $$Y_{icb} = \beta_w \cdot HS_{cb}^{age\,W} + \beta \cdot HS_{cb}^{age\,4} + \gamma \cdot C_{icb} + \theta_c + \tau_b + \varepsilon_{icb}.$$

We estimate models such as [H1] where we include our regressor of interest (β_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 H1, 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 individuals 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 H1. 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.

Robustness to using exposure variation to SFRs only: Our approach to estimating isolating the causal effect of school spending on student outcomes is somewhat complicated. While our approach makes the best use of all plausible exogenous variation in school spending due to the passage of a court-ordered school finance reform, it is helpful that our results are robust to using a simpler approach. Specifically, one may worry that our dosage predictor ($\widehat{dose}_d$) may be biased, and that because it is estimated in a first stage, we may understate the underlying noise in our final estimates. To address both these concerns directly, we estimate models that only use variation in SFR exposure for identification *and do not use any variation due to dosage*. Specifically, we use the within-county DiD variation in Head Start spending ($HS_{icb}^{age\,4}$), instrument for the natural log of public K12 spending, ppe_{idb}^{5-17}, with ($SFRExp_{idb}$), and instrument for, INT_{idb}, the interaction between Head Start and K12 spending, with ($HS_{icb}^{age\,4} \times SFRExp_{idb}$). In words, our excluded instruments are <u>two-way</u> interactions between the number of school-age years of exposure to a court-ordered SFR and Head Start spending per four-year-old when the individual was age 4. 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 [6], where $\widehat{ppe}_{idb}^{5-17}$ and $\widehat{INT}_{idb}$ are fitted values from first-stage regressions.[43] Note that this is no longer an overidentified model, but is a just identified model.

$$[6] \qquad Y_{icb} = \beta_{HS} \cdot HS_{cb}^{age\,4} + \beta_{k12} \cdot \widehat{ppe}_{idb}^{5-17} + \beta_{int} \cdot (\widehat{INT}_{idb}) + \gamma \cdot C_{icb} + \theta_d + \tau_{bdi} + \varepsilon_{idb}.$$

The results from this just identified model are presented in Appendix Table H2. As one can see, for our two main outcomes (years of education and wages), the results are very similar to those of the DiD-2SLS models. Also, formal statistical tests fail to reject that these models are the same. Importantly, (even though the standard errors are considerably larger in the simple model) our main findings are robust to this more parsimonious model. This suggests that (a) our measure of dosage is not biased and, (b) our estimated effects are robust to using approaches that do not have multiple first stages.

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

[43] Where $\hat{X}_1 = \widehat{ppe}_{idb}^{5-17}$ and $\hat{X}_2 = \widehat{INT}_{idb}$, and $w \in \{1,2\}$,
$\hat{X}_w = \pi_{w2}(SFRExp_{idb}) + \pi_{w4}(SFRExp_{idb}) \cdot HS_{cb}^{age\,4} + \gamma_w C_{idb} + \theta_{wd} + \tau_{wb}$.

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 H3).

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 influence outcomes. We achieve this by augmenting [6] and [7] to include sibling fixed effects. As one can see in Table H4, while we lose considerable precision, 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. These sibling tests also address 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 our preferred DiD-2SLS 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. NOTE: This does not exclude movers; we exclude the 3% of our sample for whom the initial address *could have been* the result of endogenous movement. The results are presented in Appendix Table H5. As one would expect, 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 (i.e. with no controls) between Head Start spending per poor 4-year old (at age 4) and instrumented ln(K12 spending) is only 0.15. To test this 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 H6. In such models, the coefficient is an economically insignificant 0.013 and the *p*-value is larger than 0.1. Taken at face value, the point estimate indicates that an exogenous 10% increase in K12 spending is associated with a mere additional $1.3 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 0.001 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 H7). 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 22.41 and 23.01 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. Looking at predicting head Start spending, the first stage F-statistic for the Head Start spending (based on rollout) is 59.17 and 60.76 in models without and with K12 variables, respectively. As such, there is a strong first stage for Head Start spending whether K12 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 28.71, conditional on Head Start spending and SFR dosage times SFR exposure. Similarly, the first stage F-statistic for Head Start rollout times SFR dosage times SFR exposure is 42.46, conditional on Head Start rollout 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 indicate that we have sufficient independent exogenous variation to credibly identify the effects of Head Start spending, the effect of K12 spending, and the effects of the interaction between the two.

Appendix Table H1:

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 (Poor children. Outcomes are measured between ages 20-45)

	Years of Education	Prob(High School Grad)	Ln(Wage)	Prob(poverty)	Prob(Ever incarcerated)
	1	2	3	4	5
$Ln(PPE_d)_{(age\ 5\text{-}17)}$	4.9251**	0.9026*	1.3588**	-0.8803**	-0.6448+
	(2.4230)	(0.5430)	(0.6351)	(0.4256)	(0.4003)
$Ln(PPE_d)_{(age\ 20\text{-}24)}$	-0.8152	0.00044	-0.1450	0.0261	0.06397
	(2.5142)	(0.4131)	(0.2805)	(0.1864)	(0.1711)
Number of person-year observations			55,706	88,124	
Number of Individuals	5,419	5,419	5,613	6,373	4,536

Robust standard errors in parentheses (clustered at childhood state 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,% black, education,% urban, population size,% voted for Strom Thurmond in 1948 Presidential election*race (proxy for segregationist preferences)) each interacted with linear cohort trends.

Table H2:

Using only School Finance Reform Exposure as Instruments for K12 Spending (Poor Children):

	1	2	3	4
	Years of Education DiD-2SLS		Ln(Wage), ages 20-50 DiD-2SLS	
Head Start Spending$_{(age\ 4)}$	0.07721***	0.0670***	0.02334***	0.0215***
	(0.01992)	(0.0189)	(0.004503)	(0.004523)
(SFR) Instrumented Ln(PPE)$_{(age\ 5\text{-}17)}$	4.0399**	9.4546**	2.0561***	2.7146***
	(1.6751)	(4.4808)	(0.4348)	(0.8493)
Head Start Spending$_{(age\ 4)}$*Instrumented ln(PPE) $_{(age\ 5\text{-}17)}$	0.6460***	0.6879***	0.1698**	0.1660**
	(0.2354)	(0.2485)	(0.06985)	(0.0705)
SFR Exposure and dosage Instruments for K12 Spending?	YES	NO	YES	NO
Only SFR Exposure Years as Instruments for K12 Spending?	--	YES	--	YES
Number of Person-year Observations	--	--	55,706	55,706
Number of Children	5,419	5,419	5,613	5,613
Number of Childhood Families	2,133	2,133	2,202	2,202
Number of School Districts	749	749	761	761
Number of Childhood Counties	600	600	610	610

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

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

Data: PSID geocode Data (1968-2015), 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 at age 4 in the county and instrumented ln(school district per-pupil spending during ages 5-17) are centered around their respective means, to facilitate interpretation of the main effects evaluated at roughly the mean; the average SFR-induced increase in school-age spending is about 10%. Results are based on 2SLS models that 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; controls for 1960 county characteristics (poverty rate,% black, education,% urban, population size, each interacted with linear cohort trends; controls for county-level per-capita gov't safety net expenditures average during childhood; and controls for childhood family characteristics (parental income/education, mother's marital status at birth, birth weight, gender), and age (cubic). The first-stage model includes as predictors the school-age years of exposure to school finance reform. Results in columns (2) & (4) DO NOT include school finance reform "dosage" intensity terms as instruments (i.e., without 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), while columns (1) & (3) do include SFR "dosage" as in preferred 2SLS-DiD presented in Table 1a. There exists a significant first-stage.

Table H3.

Examining Exogeneity of Head Start and K-12 Spending (Poor 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$_{(\text{age } 4)}$/1000	-0.0044313	-0.0037102	0.0002552	-0.000113
	(0.0046289)	(0.0048768)	(0.0007137)	(0.0007671)
Ln(K12 Per-pupil Spending)$_{(\text{age } 5\text{-}17)}$	0.7713586	0.7905367	0.0432606	0.0438955
	(0.8242292)	(0.8474821)	(0.1618254)	(0.1904868)

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

Data: PSID geocode Data (1968-2015), matched with childhood school and neighborhood characteristics. Analysis sample includes PSID individuals born 1950-1976, followed into adulthood through 2015. 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, 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. The results presented in this Table show that, holding constant school district fixed effects and birth year fixed effects, identifying variation in Head Start spending increases are NOT significantly related to (changes in) childhood family/community SES characteristics. 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.

Table H4: *Within Family Model: 2SLS-Difference-in-Difference Estimates of Early and K12 education Spending on Adult Outcomes (Poor Children)*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Prob(High School Grad)		Years of Completed Education		Ln(Wage), ages 20-50		Annual Incidence of Poverty, age 20-50		Prob(Ever Incarcerated)	
	DiD-2SLS	2SLS-IV	DiD-2SLS	2SLS-IV	DiD-2SLS	2SLS-IV	DiD-2SLS	2SLS-IV	DiD-2SLS	2SLS-IV
Head Start Spending$_{\text{(age 4)}}$	0.02082*	0.08133**	0.1392***	0.2519*	0.01269**	0.0708*	-0.004733+	-0.02100	-0.0206***	-0.0281
	(0.01252)	(0.04115)	(0.03635)	(0.1525)	(0.005874)	(0.0426)	(0.002972)	(0.02398)	(0.0051)	(0.0343)
(SFR) Instrumented Ln(PPE)$_{\text{(age 5-17)}}$	1.5534**	1.2292***	6.0377***	4.5989**	0.9450*	1.3371***	-0.5449*	-0.9089***	0.2587	-0.1613
	(0.7661)	(0.4381)	(2.2362)	(2.0073)	(0.5000)	(0.4016)	(0.3273)	(0.3334)	(0.2720)	(0.4227)
Head Start Spending$_{\text{(age 4)}}$*ln(PPE)$_{\text{(age 5-17)}}$	0.1062*	0.2374**	0.8159***	0.9395**	0.06292	0.07938	-0.06652+	-0.2153***	0.0744	-0.0360
	(0.05974)	(0.1047)	(0.2958)	(0.4759)	(0.07046)	(0.1091)	(0.04744)	(0.06991)	(0.0462)	(0.1047)
Number of Person-year Observations	5419	5419	5419	5419	55706	55706	88124	88124	4536	4536
Number of Families	2133	2133	2133	2133	2202	2202	2301	2301	1727	1727

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

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

Data: PSID geocode Data (1968-2015), matched with childhood school and neighborhood characteristics. Sample includes all individuals born 1950-1976 whose parents were in the bottom quartile of the income distribution, and who have been followed into adulthood.

Models: (non-Instrumented & Instrumented) Head Start Spending per poor 4-year old at age 4 in the county and instrumented ln(school district per-pupil spending during ages 5-17) are centered around their respective means, to facilitate interpretation of the main effects evaluated at roughly the mean; the average SFR-induced increase in school-age spending is about 10%. Results are based on 2SLS-IV models that include: childhood family 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; controls for 1960 county characteristics (poverty rate,% black, education,% urban, population size, each interacted with linear cohort trends; controls for county-level per-capita gov't safety net expenditures average during childhood; and controls for child-specific family characteristics (parental income, mother's marital status at birth, birth weight, gender), and age (cubic). The first-stage model includes 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 instrument used for Head Start spending per poor 4-year old is an indicator for whether there was any Head Start center in the county at age 4 (based on the program's rollout timing variation only). There exists a significant first-stage.

Table H5:

Early Address Sample:

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

	Children from Poor Households				
	1	2	3	4	5
	Years of Education	Prob(High School Grad)	Ln(Wage), ages 20-50	Annual Incidence of Poverty, ages 20-50	Prob(Incarcerati on)
Head Start Spending$_{(age\ 4)}$/1000	0.06754***	0.01420**	0.02198***	-0.01649***	-0.005557*
	(0.01584)	(0.006972)	(0.004307)	(0.004463)	(0.003232)
Ln(K12 Per-pupil Spending)$_{(age\ 5\text{-}17)}$	4.9596***	1.0533**	2.3668***	-0.8955**	-0.7171**
	(1.8986)	(0.4259)	(0.5023)	(0.3964)	(0.3188)
Interaction	0.6981***	0.1316**	0.1749***	-0.1036***	-0.05540*
	(0.2470)	(0.06662)	(0.06136)	(0.03451)	(0.03313)
Number of Person-year Observations	--		53,970	84,326	--
Number of Children	5,071	5,071	5,280	5,971	4,408

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

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

Data: PSID geocode Data (1968-2015), 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, who have been followed into adulthood, and for whom earliest available address predates Head Start rollout and school finance reform.

Models: Head Start Spending per poor 4-year old at age 4 in the county and instrumented ln(school district per-pupil spending during ages 5-17) are centered around their respective means, to facilitate interpretation of the main effects evaluated at roughly the mean; the average SFR-induced increase in school-age spending is about 10%. Results are based on 2SLS-IV models that 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; controls for 1960 county characteristics (poverty rate,% black, education,% urban, population size, each interacted with linear cohort trends; controls for county-level per-capita gov't safety net expenditures average during childhood; and controls for childhood family characteristics (parental income/education, mother's marital status at birth, birth weight, gender), and age (cubic). The first-stage model includes 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.

Table H6:

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 $\text{old}_{(\text{age } 4)}$ (in 000s)	(SFR) Instrumented Ln(School District Per-pupil $\text{Spending})_{(\text{age } 5\text{-}17)}$
County Head Start Spending per Poor 4-year $\text{old}_{(\text{age } 4)}$ (in 000s)	--	0.0001633
		(0.0003878)
(SFR) Instrumented Ln(School District Per-pupil $\text{Spending})_{(\text{age } 5\text{-}17)}$	0.0133935	--
	(2.853857)	
Number of Children	5,419	5,419

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 0.15; 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,% black, education,% urban, population size,% 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 includes 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 H7:

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

	(1)	(2)	(3)	(4)	(5)	(6)
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 and SFR dosage*SFR exposure*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 and SFR dosage*SFR exposure*Head Start 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	22.41	23.01	28.71	42.46	59.17	60.76

Robust standard errors clustered at childhood state level.

Data: PSID geocode Data (1968-2015), 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 its mean 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: 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, roll-out 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,% black, education,% urban, population size) 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 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 instrument used for Head Start spending per poor 4-year old is an indicator for whether there was any Head Start center in the county at age 4 (based on the program's rollout timing variation only); and in column (6) this instrument is interacted with school-age years of exposure to school finance reform*dosage (reform-induced change in school spending based on the timing and type of court-ordered reform).

Appendix Table H8. *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: Head Start Spending per Poor 4-year old$_{(age\ 4)}$ (in 000s)	0.07640***	0.07742***	0.1077**	0.07392***	0.07000***	0.08065***	0.07577***	0.07830***	0.08185***	0.07734***
	(0.01967)	(0.01942)	(0.04214)	(0.02337)	(0.01599)	(0.02726)	(0.01961)	(0.02009)	(0.01997)	(0.01981)
Age1: Head Start Spending per Poor 4-year old$_{(age\ 1)}$ (in 000s)		0.01329								
		(0.03527)								
Age2: Head Start Spending per Poor 4-year old$_{(age\ 2)}$ (in 000s)			-0.07522							
			(0.05727)							
Age3 Head Start Spending per Poor 4-year old$_{(age\ 3)}$ (in 000s)				0.01079						
				(0.02878)						
Age5: Head Start Spending per Poor 4-year old$_{(age\ 5)}$ (in 000s)					0.02417					
					(0.03274)					
Age6: Head Start Spending per Poor 4-year old$_{(age\ 6)}$ (in 000s)						-0.01763				
						(0.05074)				
Age7: Head Start Spending per Poor 4-year old$_{(age\ 7)}$ (in 000s)							0.005178			
							(0.02362)			
Age8: Head Start Spending per Poor 4-year old$_{(age\ 8)}$ (in 000s)								-0.009519		
								(0.02046)		
Age9: Head Start Spending per Poor 4-year old$_{(age\ 9)}$ (in 000s)									-0.05718**	
									(0.02868)	
Age10: Head Start Spending per Poor 4-year old$_{(age\ 10)}$ (in										-0.008806
										(0.01934)
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

*** 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,% black, education,% urban, population size,% 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 Table H9. *Placebo Tests: Effects of Head Start Access by Child Age on Educational Attainment, Low-Income*

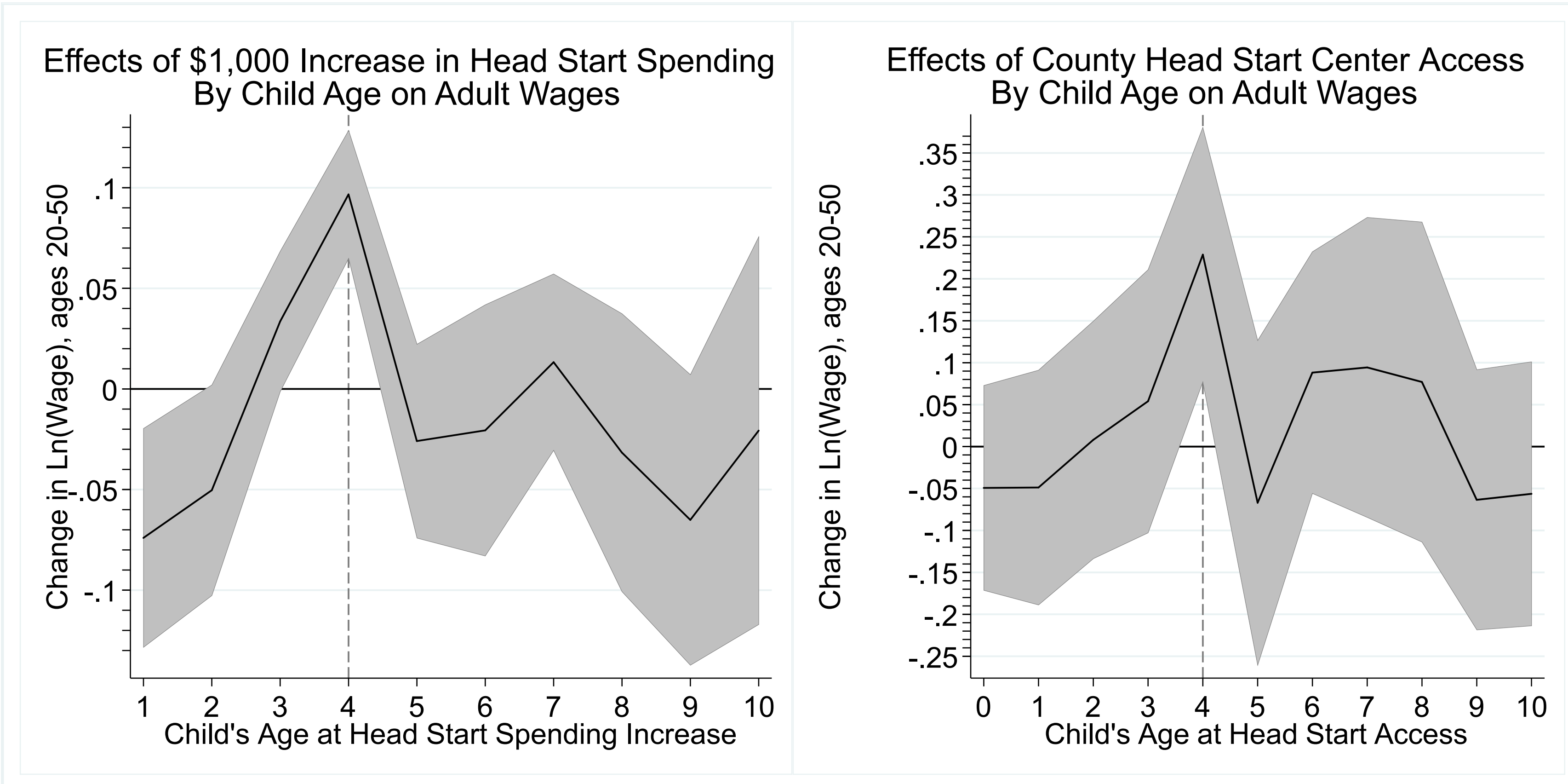
	Completed Years of Education									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Age 4: Head Start Center	0.8889**	0.9009**	0.8228*	1.1938**	1.2516**	1.0180**	0.7025*	0.7774**	0.7791**	0.7402**
	(0.3710)	(0.4065)	(0.4972)	(0.5271)	(0.5001)	(0.4340)	(0.4025)	(0.3552)	(0.3599)	(0.3363)
Age1: Head Start Center		-0.02754								
		(0.2867)								
Age2: Head Start Center			0.08828							
			(0.4110)							
Age3 Head Start Center				-0.4204						
				(0.3922)						
Age5: Head Start Center					-0.4697					
					(0.4667)					
Age6: Head Start Center						-0.2205				
						(0.4303)				
Age7: Head Start Center							0.3626			
							(0.3297)			
Age8: Head Start Center								0.2284		
								(0.3684)		
Age9: Head Start Center									0.2445	
									(0.4334)	
Age10: Head Start Center										0.4148
										(0.4142)
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	761	761	761	761	761	761	761	761	761	761
Number of Childhood	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,% black, education,% urban, population size,% 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).

Figure H1: *Effect of Head Start Spending and Rollout on Adult Wages by Age*



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 a model with no other ages included.

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
	Prob(High School Grad)	Years of Education	Ln(Wage), 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)	-0.0233	0.0118	0.0024	-0.0033	-0.0021
	(0.0206)	(0.0690)	(0.0269)	(0.0172)	(0.0053)

Robust standard errors in parentheses (clustered at childhood state level)
*** p<0.01, ** p<0.05, * p<0.1
Data: PSID geocode Data (1968-2015), 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: 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, roll-out 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 yrs)), timing of state-funded Kindergarten intro and timing of tax limit policies; controls for 1960 county characteristics (poverty rate,% black, education,% urban, population size) 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).

Appendix J

Table J1: *Poor vs Non-Poor Children (DiD-2SLS Models)*

	(1)	(2)	(3)	(4)	(5)	(6)
	Years of Education			Ln(Wage), ages 20-50		
	Poor	Non-Poor	Difference	Poor	Non-Poor	Difference
Head Start Spending$_{(age\ 4)}$	0.07721***	0.008866	0.0683***	0.02334***	0.006901	0.0164**
	(0.01992)	(0.01635)	(0.0258)	(0.004503)	(0.005408)	(0.0070)
(SFR) Instrumented Ln(PPE)$_{(age\ 5\text{-}17)}$	4.0399**	2.4192**	1.6207	2.0561***	0.7351**	1.3210**
	(1.6751)	(1.1645)	(2.0401)	(0.4348)	(0.3035)	(0.5302)
Head Start Spending$_{(age\ 4)}$*ln(PPE)$_{(age\ 5\text{-}17)}$	0.6460***	0.02972	0.6163**	0.1698**	0.02577	0.1440*
	(0.2354)	(0.1937)	(0.3048)	(0.06985)	(0.03090)	(0.0764)
Number of Person-year Observations	--	--		55,706	90,771	
Number of Children	5,419	7,983		5,613	8,195	
Number of Childhood Families	2,133	3,530		2,202	3,593	
Number of School Districts	749	1,156		761	1,169	
Number of Childhood Counties	600	891		610	908	

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

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

Data: PSID geocode Data (1968-2015), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 who have been followed into adulthood (218,594 person-year observations; 15,232 individuals; 4,990 childhood families; 1,427 school districts; 1,120 childhood counties).

Models: (non-Instrumented & Instrumented) Head Start Spending per poor 4-year old at age 4 in the county and instrumented ln(school district per-pupil spending during ages 5-17) are centered around their respective means, to facilitate interpretation of the main effects evaluated at roughly the mean; the average SFR-induced increase in school-age spending is about 10%. Results are based on 2SLS-IV models that 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; controls for 1960 county characteristics (poverty rate,% black, education,% urban, population size, each interacted with linear cohort trends; controls for county-level per-capita gov't safety net expenditures average during childhood; and controls for childhood family characteristics (parental income/education, mother's marital status at birth, birth weight, gender), and age (cubic). The first-stage model includes 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 instrument used for Head Start spending per poor 4-year old is an indicator for whether there was any Head Start center in the county at age 4 (based on the program's rollout timing variation only). There exists a significant first-stage. The marginal effects related to Head Start access are based on the average county Head Start spending when there is a center (~$4,230 (in real 2000 dollars))--i.e., marginal effects are evaluated for roughly a $4K increase in Head Start spending (to contrast impact of having access vs no access to Head Start center).

Appendix K:

Table K1: *K12 Spending Effects by Age at First School Finance Reform*

Dependent Variable:	Years of Education	Ln(Wage), ages 20-50	Years of Education	Ln(Wage), ages 20-50
	Exposed to SFR 1-10 years: (after 7)		Exposed to SFR 11+ years: (before 7)	
(SFR) Instrumented Ln(PPE)$_{(age\ 5\text{-}17)}$	4.3942*	2.1295**	2.1434	0.8305**
	(2.3134)	(0.9008)	(2.0695)	(0.3604)
Head Start Spending$_{(age\ 4)}$*ln(PPE)$_{(age\ 5\text{-}17)}$	0.1900	0.3313*	0.6331**	0.1338**
	(0.3610)	(0.1774)	(0.2591)	(0.0630)
	Exposed to SFR 1-9 years: (after 8)		Exposed to SFR 10+ years: (before 8)	
(SFR) Instrumented Ln(PPE)$_{(age\ 5\text{-}17)}$	7.0847**	2.6364***	4.5313***	1.2779***
	(3.2518)	(0.9405)	(1.1263)	(0.3151)
Head Start Spending$_{(age\ 4)}$*ln(PPE)$_{(age\ 5\text{-}17)}$	0.3217	0.3512	0.6683**	0.1423**
	(0.4321)	(0.2376)	(0.2745)	(0.0662)
	Exposed to SFR 1-8 years: (after 9)		Exposed to SFR 9+ years: (before 9)	
(SFR) Instrumented Ln(PPE)$_{(age\ 5\text{-}17)}$	6.3156+	1.8053**	3.8329***	1.3591***
	(4.8084)	(0.7514)	(1.1443)	(0.4269)
Head Start Spending$_{(age\ 4)}$*ln(PPE)$_{(age\ 5\text{-}17)}$	0.0590	0.1491	0.6054***	0.1553**
	(0.8043)	(0.1113)	(0.2230)	(0.0665)

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

Data: PSID geocode Data (1968-2015), 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: Each panel is based on a single regression in which all the K12 instruments and K12 spending variables are interacted with indicator variables connoting whether an SFR occurred in their childhood state before a particular age or after a particular age (top panel is 7, middle panel is 8, and the bottom panel I 9 years old). 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, and both the county per-capita Medicaid and food stamps spending variables are also included and centered around their respective means, 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: 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, roll-out 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), 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; and each of these variables interacted with an indicator for whether individual was exposed to an SFR before or after 7 (top panel), 8 (middle panel), or 9 (bottom panel) years old. There exists a significant first-stage for all variables in all models.

Table K2: *First Stage Estimates*

	1	2	3	4	5
Endogenous Variable (Dependent Variable of First Stage)	$Ln(PPE)_{(age\ 5\text{-}17)}$		Head Start Spending*Ln(PPE)		Head Start Spending
Model	DID-2SLS	2SLS-2SLS	DID-2SLS	2SLS-2SLS	2SLS-2SLS
Head Start Spending Instruments					
Head Start $Exposure_{(age4)}$		-0.0487		0.3228+	3.0333***
		(0.0516)		(0.2088)	(0.3891)
K12 Spending Instruments					
# of School-age years of SFR exposure	0.0091*	0.0291***	0.0387+	-0.0860+	0.0830
	(0.0048)	(0.0086)	(0.0261)	(0.0629)	(0.1217)
(School-age years of SFR exposure)*(Dosage quartile2)	0.0049	0.0041	-0.0165	-0.1242**	-0.2763***
	(0.0047)	(0.0076)	(0.0273)	(0.0537)	(0.0820)
(School-age years of SFR exposure)*(Dosage quartile3)	0.0053+	0.0032	-0.0781***	-0.0567***	-0.0080
	(0.0035)	(0.0054)	(0.0153)	(0.0204)	(0.0831)
(School-age years of SFR exposure)*(Dosage top quartile)	0.0110***	-0.0454***	-0.0819***	0.1548***	-0.0592
	(0.0033)	(0.0084)	(0.0261)	(0.0314)	(0.1476)
Instruments for Interaction (DID-2SLS models)					
# of School-age years of SFR exposure*Head Start $spending_{(age4)}$	-0.0009		0.0155***		
	(0.0007)		(0.0034)		
(School-age years of SFR exposure)*(Dosage quartile2)*Head Start	-0.0068***		0.0294***		
	(0.0015)		(0.0098)		
(School-age years of SFR exposure)*(Dosage quartile3)*Head Start	0.0015		-0.0136**		
	(0.0014)		(0.0057)		
(School-age years of SFR exposure)*(Dosage top quartile)*Head Start	0.0025		-0.0160+		
	(0.0017)		(0.0109)		
Instruments for Interaction (2SLS-2SLS models)					
# of School-age years of SFR exposure*Head Start $Exposure_{(age4)}$		-0.0249***		0.1390***	-0.0156
		(0.0048)		(0.0284)	(0.0399)
(School-age years of SFR exposure)*(Dosage top quartile)*Head Start		0.0614***		-0.2423***	0.0429
		(0.0060)		(0.0205)	(0.0834)
Number of Children	5,419	5,419	5,419	5,419	5,419

Robust standard errors in parentheses (clustered at childhood state level)
*** p<0.01, ** p<0.05, * p<0.1
Data: PSID geocode Data (1968-2015), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born

Models: The set of controls 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, roll-out 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) 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 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 instrument used for Head Start spending per poor 4-year old is an indictor for whether there was any Head Start center in the county at age 4 (based on the program's rollout timing variation only); and this instrument is interacted with school-age years of exposure to school finance reform*dosage (reform-induced change in school spending based on the timing and type of court-ordered reform). There exists a strong first-stage in all models.

Table K3: *2SLS-DiD Estimates With and Without Controls for Other Policies: Poor Children*

	1	2	3	4
	Years of Education		Ln(Wage), ages 20-50	
	Parsimonious	Full Set of Controls	Parsimonious	Full Set of Controls
Head Start Spending$_{(age\ 4)}$	0.0751***	0.07721***	0.0226***	0.02334***
	(0.0221)	(0.01992)	(0.0042)	(0.004503)
Ln(PPE)$_{(age\ 5\text{-}17)}$	3.4683**	4.0399**	1.6507***	2.0561***
	(1.6641)	(1.6751)	(0.5387)	(0.4348)
Head Start Spending$_{(age\ 4)}$*ln(PPE)$_{(age\ 5\text{-}17)}$	0.6227***	0.6460***	0.1787***	0.1698**
	(0.2208)	(0.2354)	(0.0656)	(0.06985)
Number of Person-year Observations	--	--	55,706	55,706
Number of Children	5,419	5,419	5,613	5,613

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

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

Data: PSID geocode Data (1968-2015), 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 at age 4 in the county and instrumented ln(school district per-pupil spending during ages 5-17) are centered around their respective means, to facilitate interpretation of the main effects evaluated at roughly the mean; the average SFR-induced increase in school-age spending is about 10%. Results are based on 2SLS-IV-DiD models. Columns (1), (3), (5) include a more parsimonious set of controls that include: school district fixed effects, race-specific year of birth fixed effects, race*census division-specific birth year trends, and controls for childhood family characteristics (parental income/education, mother's marital status at birth, birth weight, gender), age (cubic). Columns (2), (4), (6) use complete set of controls that in addition to the aforementioned include: controls at the county-level for the timing of school desegregation*race, hospital desegregation*race; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, each interacted with linear cohort trends; controls for county-level per-capita gov't safety net expenditures average during childhood. The first-stage model includes 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.

Table K4: *OLS Estimates of Interactive Effects of Head Start and K12 Spending (Poor Children)*

	1	2	3	4	5
	Years of Education	Prob(High School Grad)	Ln(Wage), ages 20-50	Annual Incidence of Poverty, ages 20-50	Prob(Ever Incarcerated)
Head Start Spending$_{(age\ 4)}$	0.07965***	0.02570***	0.01327***	-0.01116***	-0.006562*
	(0.01271)	(0.006439)	(0.004868)	(0.002061)	(0.003698)
Ln(PPE)$_{(age\ 5-17)}$	0.6542	0.2283	0.2537**	-0.07677	-0.05556
	(0.5300)	(0.1823)	(0.1228)	(0.06654)	(0.07668)
Head Start Center$_{(age\ 4)}$*ln(PPE)$_{(age\ 5-17)}$	0.2473**	0.04139	0.04435**	-0.02757**	-0.004091
	(0.1210)	(0.02703)	(0.02135)	(0.01063)	(0.01583)
Number of Person-year Observations	--	--	55,706	88,124	--
Number of Children	5,419	5,419	5,613	6,373	4,536

Robust standard errors in parentheses

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

Data: PSID geocode Data (1968-2015), 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, who have been followed into adulthood.

Models: Head Start Spending per poor 4-year old at age 4 in the county and ln(school district per-pupil spending during ages 5-17) are centered around their respective means, to facilitate interpretation of the main effects evaluated at roughly the mean. Results are based on OLS models that 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; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, each interacted with linear cohort trends; controls for county-level per-capita gov't safety net expenditures average during childhood; and controls for childhood family characteristics (parental income/education, mother's marital status at birth, birth weight, gender), and age (cubic).

Table K5: *OLS Estimates of Interactive Effects of Head Start and K12 Spending (Non-Poor Children)*

	1	2	3	4	5
	Years of Education	Prob(High School Grad)	Ln(Wage), ages 20-50	Annual Incidence of Poverty, ages 20-50	Prob(Ever Incarcerated)
Head Start Spending$_{(age\ 4)}$	0.01671	0.00094	0.006004	-0.00088	-0.000725
	(0.01845)	(0.003676)	(0.004257)	(0.001115)	(0.001451)
Ln(PPE)$_{(age\ 5\text{-}17)}$	-0.2178	0.01140	0.02731	-0.009794	0.05600
	(0.2585)	(0.06029)	(0.08434)	(0.01899)	(0.03661)
Head Start Center$_{(age\ 4)}$*ln(PPE)$_{(age\ 5\text{-}17)}$	-0.07470	-0.002712	0.01167	-0.001164	0.007585
	(0.05749)	(0.01116)	(0.01642)	(0.003119)	(0.004769)
Number of Person-year Observations	--	--	90,771	130,470	--
Number of Children	7,983	7,983	8,195	8,859	5,140

Robust standard errors in parentheses

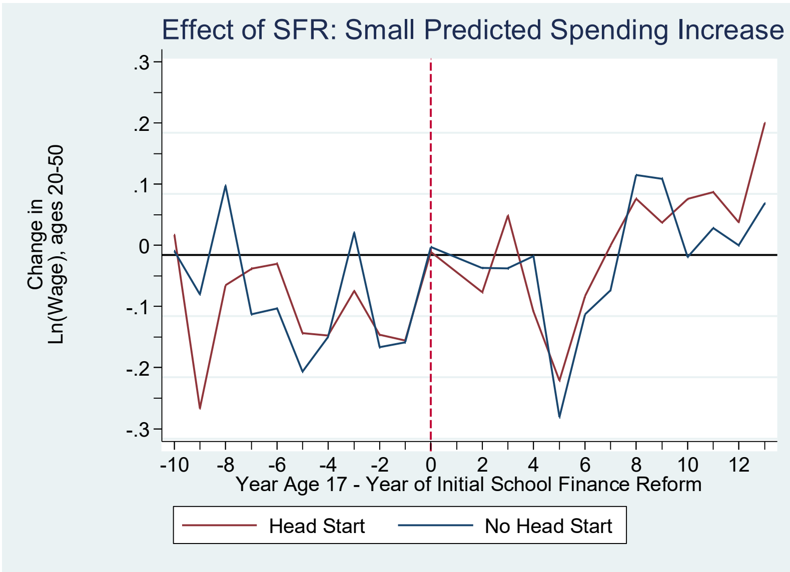
*** p<0.01, ** p<0.05, * p<0.1

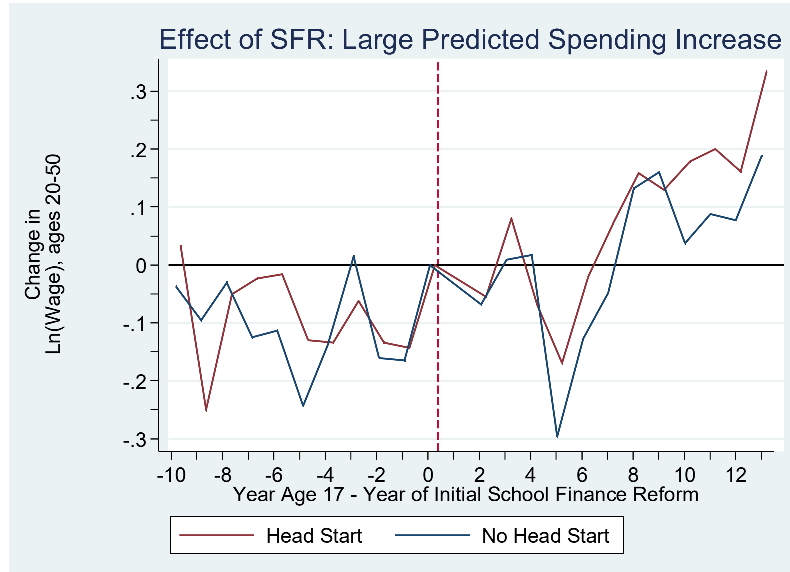
Data: PSID geocode Data (1968-2015), matched with childhood school and neighborhood characteristics. Analysis sample includes all PSID individuals born 1950-1976 whose parents were NOT in the bottom quartile of the income distribution, who have been followed into adulthood.

Models: Head Start Spending per poor 4-year old at age 4 in the county and ln(school district per-pupil spending during ages 5-17) are centered around their respective means, to facilitate interpretation of the main effects evaluated at roughly the mean. Results are based on OLS models that 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; controls for 1960 county characteristics (poverty rate, percent black, education, percent urban, population size, each interacted with linear cohort trends; controls for county-level per-capita gov't safety net expenditures average during childhood; and controls for childhood family characteristics (parental income/education, mother's marital status at birth, birth weight, gender), and age (cubic).

Appendix L

Figure L1: *Interaction Event Study for Wages*





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 (blue line), and those who were exposed to any county Head Start spending at age 4 (red line), to highlight the role of dynamic complementarity. Transitory fluctuations in adult wages may introduce more noise to these event study figures than comparable ones for educational attainment (as presented in Figure 5). The event study models include: school district fixed effects, race-specific year of birth fixed effects, age (cubic), controls for childhood family characteristics (parental income/education, mother's marital status at birth, birth weight, gender) and same set of other controls as main models.

Appendix M

Estimating Head Start Participation Rates

The ratio of enrolled students to the income-eligible age-eligible population in a given year *is not* the same as a specific cohort's participation rate by kindergarten entry.

To relate these enrollments to participation rates at the individual child level, for each kindergarten entry cohort we computed the cumulative likelihood across all age-eligible years that an income-eligible child would enroll in Head Start.

To illustrate this point, suppose for simplicity that Head Start had only the summer program. For example, the annual enrollment rate in summer programs was about 22 percent between 1965 and 1967. The cohort of income-eligible children entering kindergarten in 1965 could only have enrolled at age 5 and would have a 22 percent participation rate. However, the cohort of income-eligible children that entered kindergarten in 1966 could have enrolled at age 4 or 5, so (assuming that participants enroll for one year and not multiple years) their cohort's participation rate by kindergarten entry would be 44 percent (i.e., the sum of the likelihood of participation during ages 4 or 5: 22 + 22). All subsequent cohorts could have enrolled at ages 3, 4, or 5, so that post-1966 cohorts' participation rate by kindergarten entry (across all age-eligible years) is the running total annual summer enrollment ratio for the three years preceding kindergarten entry. Similarly, assuming that participants enroll for one year and not multiple years, a specific cohort's full-year participation rate by kindergarten entry is the running total annual full-year enrollment ratio for the two years preceding kindergarten entry.

To avoid double-counting individuals who enrolled in both the summer program and the full-year programs, we assume that 40 percent of full-year enrollees were previously in a summer program.

Appendix N

Details of the back of the envelope calculations:

It is helpful to define some parameters. The proportion of poor children in a county is p. The average per-student cost of rolling out the average Head Start center is the cost of increasing Head Start spending per poor-4-year old by $4,320. The average cost of this increase is simply $4320*p$. The marginal effect of rolling out the average Head Start center for a county (π_{HS}), is a poverty-weighted average of the effect of a $4,320 increase in Head Start spending on low-income children ($\delta_{HS,\,poor}$), and that for non-poor children ($\delta_{HS,non}$). Because Head Start has no effect on non-poor children, this simplifies to [8] below.

[8] $$\pi_{HS} = p\delta_{HS,poor}.$$

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 $4,320 in Head Start spending per poor-4-year old. During our sample period, K12 spending was roughly $4,000 per student per year *on average*. Assuming a 7% 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 4320p/34000=(p*12.7)% increase in K12 spending. We define $\delta_{K12,poor}$ and $\delta_{K12,non}$ as the effect of increasing K12 spending by one% 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] $$\pi_{K12} = (p\delta_{K12,poor} + (1-p)\delta_{K12,non})(12.7p)$$

The ratio shown in [10] between these two equations π_{HS}/π_{K12} is the relative effectiveness of rolling out Head Start (from having no center) and spending the same amount across all children from that same cohort in the county in the K12 system.

[10] $$\frac{\pi_{HS}}{\pi_{K12}} = \frac{p\delta_{HS,poor}}{(p\delta_{K12,poor}+(1-p)\delta_{K12,non})(12.7\text{p})} = \frac{\delta_{HS,poor}/(12.7)}{p(\delta_{K12,poor}-\delta_{K12,non})+\delta_{K12,non}}$$

The relative marginal effect of Head Start rollout and the equivalent spending in the K12 system 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.

Appendix O

Additional Contextual Details on Head Start

- Children who are 4 years old and whose family income is below the federal poverty guidelines (or is on public assistance programs AFDC or SSI) are eligible for the program. Beginning in 1972 (as part of the Economic Opportunity Act Amendment) at least 10% of children per center must have a disability (irrespective of the family income 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).

- 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.