The Spillover Effects of Universal Pre-K: Modern Evidence on Mothers' Labor Supply Responses

Elise A. Marifian*

October 31, 2021

Abstract

This paper provides new empirical evidence of the effect of universal pre-k on mothers' labor supply. I study the 2014 introduction of Pre-K For All in New York City, which increased the city's capacity of publicly-funded, full-day pre-K seats from 20,000 in 2013 to approximately 70,000 in 2015. To identify the causal impact on mothers' labor supply, I use a difference-in-differences (DID) research design that leverages age and residency eligibility requirements to construct two different comparison groups of mothers: mothers of same-aged children in nearby New Jersey, and mothers of slightly older children that reside in NYC. I find that Pre-K For All induces mothers to increase their labor market activity both on the extensive and intensive margins, with participation rates increasing between 3.3 and 8.4 percentage points and usual hours worked per week increasing between 0.7 to 2.4 hours, depending on the time frame considered. These results suggest that access and costs of child care may be an important factor reducing the labor supply of mothers with preschool-aged children.

*Department of Economics and Institute for Research on Poverty, University of Wisconsin, Madison. I gratefully acknowledge financial support from the Institute for Education Sciences and the Juli Plant Grainger Summer Fellowship at the Department of Economics, University of Wisconsin, Madison. I am indebted to Jeff Smith, Chris Taber, and Jesse Gregory for their guidance and comments on earlier drafts. I also appreciate feedback from Matt Wiswall, Corina Mommaerts, Felix Elwert, Jonathan Becker, Sandra Spirovska, Elan Segarra, Amrita Kulka, Lois Miller, Natalia Serna, Erika Frost, and Joanna Venator. All errors are my own.

1 Introduction

A large literature has documented the benefits of high quality early childhood education, especially for economically disadvantaged students.[1] This evidence, combined with the fact that a sizable share of 4 year olds do not enroll in pre-primary programs[2], has motivated policymakers' recent enthusiasm for broader preschool access and has spurred a range of policies to increase preschool enrollments[3], including universal publicly funded preschool policies.[4] Advocates argue that improving the accessibility and affordability of high quality preschool will increase children's kindergarten readiness and reduce socioeconomic and racial disparities in school preparedness and academic achievement, with possibly lasting effects on children's long-run economic prospects. Educational production, however, is not the only process that may change as preschool enrollments increase. Subsidizing preschool increases the relative price of alternative sources of schooling and childcare, which in turn may affect parents' labor market incentives and educational consumption decisions. Maternal labor supply in particular may be sensitive to free preschool, especially in places where market-based childcare accounts for a larger share of family income.

My objective in this paper is to quantify the spillover effects of increased preschool access on mothers' labor market behaviors. Focusing on New York City's September 2014 introduction of "Pre-K For All" (UPK)—a universal pre-Kindergarten policy that offers free, full-day pre-K to all 4 year old residents—I estimate the causal effect of introducing universal pre-K on the labor force participation and hours worked of eligible mothers. Given the rapid implementation of the policy and the clear geographic- and age-eligibility restrictions, this empirical setting lends itself to a difference-in-differences (DID) research design. These requirements give rise to two contrasts that can be leveraged to estimate various DID models, as well as a "triple differences" (DDD) model that combines the contrasts into a single model. Specifically, I construct two comparison groups against which to measure the effect for treated mothers: (1) Mothers of 4 year olds who reside outside New York City's geographic boundary but within the NYC metropolitan area, and NYC mothers of slightly older children.

Quantifying mothers' responses to preschool and related childcare policies can have important implications for the design of a broad range of labor-related policies, including

[1] Two relevant sources are Heckman (2006) and Duncan et al. (2007).

[2] In 2016, 42 percent of 3 year olds and 66 percent of 4 year olds were enrolled in pre-primary programs in the U.S. (National Center for Education Statistics, U.S. Department of Education).

[3] According to the National Institute for Early Education Research (NIEER) of Rutgers University, the last decade has seen an increase in city-based initiatives to improve 4 year old pre-K access and quality, with 20 out of the nation's 40 largest cities now raising local funds to support these objectives (Voyles and Ruess (2019)).

[4] Currently, only three states offer universal pre-K: Georgia, Oklahoma, and Florida. (See Fitzpatrick (2010) for additional details on the Georgia and Oklahoma programs). Some large cities have introduced universal pre-K in the last decade, the most well-known being Washington, D.C. in 2009 and New York City in 2014, while Chicago and Boston recently announced plans to offer universal pre-K within four to five years. See https://www.educationdive.com/news/more-cities-implement-universal-pre-k-when-state-national-efforts-fall-sho/528273/ and https://www.boston.gov/news/15-million-investment-universal-pre-k-guarantee-equitable-access-free-high-quality-pre-k-all.

minimum wage laws, tax credits and subsidies, and paid family leave. Of course, universal pre-K is just one of many childcare policies that can affect mothers' labor supply decisions, but it is beyond the scope of this paper to conduct a comprehensive assessment of how these responses vary across childcare policy instruments and ages.[5] Nevertheless, given the sheer size of the New York City preschool market, and the fact that the rapidly-implemented Pre-K For All policy is regarded as a model for other cities interested in universal pre-K, estimation of the treatment effect in this setting has clear policy relevance. To preview the results, I find that NYC's expansion to universal pre-K causes an economically significant increase in mothers' labor force participation, with estimates ranging from 3.3 to 12.7 percentage points depending on which post policy years are used. I also find a smaller effect on usual hours worked per week, which increases by approximately 2.4 hours in the first year of the policy, and 0.7 to 1.8 hours over the 2014-2016 period.

A large literature in economics has examined how public policies affect mothers' labor supply.[6] Within this broad topic, various papers have focused on quantifying the effect of public schooling on mothers' labor supply, with Gelbach (2002), Cascio (2009), Fitzpatrick (2010), Fitzpatrick (2012), and Herbst (2017) among the more related papers that leverage quasi-experimental variation. Gelbach (2002) is one of the earlier papers to examine public schooling's effect on maternal labor supply. Although public kindergarten becomes available to all families once their child is age eligible, enrollment is potentially endogenous because some parents choose to delay sending their child to school until the following year, while others choose to enroll in private school. Gelbach addresses this issue using a child's quarter of birth (QOB) as an instrument for public kindergarten enrollment, since children face differential public school eligibility depending on when they were born and their local age cutoff (which varies between July 1 and Dec. 31).[7] His two-stage least squares (2SLS) estimates using 1980 Census data indicate strong and statistically significant positive effects of public school enrollment on the employment (4 percentage points), weeks worked (3.60 weeks), usual hours (2.23 hours), and hours worked last week (2.71 hours) of single mothers whose youngest child is five years old.[8]

Within the literature examining the public schooling of young children and maternal labor force participation, the study most similar to mine in terms of research design is Cascio (2009). Cascio uses a difference-in-differences design to examine how maternal labor supply

[5] Olivetti and Petrongolo (2017) explore how family policies differ across countries and their effects on economic outcomes, particularly for women. Furthermore, a number of studies have explored the policies that promote labor force participation of women in contexts outside the U.S. Cascio et al. (2015) offer a synthesis of this literature's results.

[6] For example, in an earlier paper using a difference-in-differences design to study this topic, Eissa and Liebman (1996) examine how the 1987 expansion of the Earned Income Tax Credit (EITC) affected single mothers' labor force participation, which theory predicts would increase. Consistent with this prediction, they estimate an increase of 2.8 percentage points.

[7] Gelbach focuses on public kindergarten but also examines preschool (3 and 4 year olds) enrollment. However, since QOB is related both to increased private and public school enrollment, his (younger children) IV estimates identify the effects of both free public preschool and reduced-price private preschool enrollments on maternal labor outcomes.

[8] Gelbach's results for married mothers are smaller in magnitude yet similar when converted to percentage effects. His OLS estimates are incorrectly signed and large in magnitude.

was affected by the introduction of kindergarten in U.S. states over the 1960s, 1970s, and 1980s. Using variation in funding across states and over time, she estimates both DID and DDD models, and in contrast to Gelbach (2002), does not detect a significant effect on maternal employment. Fitzpatrick (2010) is most similar to my paper in terms of the age and policy studied, as she approaches the topic focusing on preschoolers, specifically the 1990s introduction of universal pre-K in Oklahoma and Georgia and its effects on preschool enrollment and mothers' employment, weeks of work, usual hours, wages, and public assistance receipt. Using restricted-access data from the 2000 Census, Fitzpatrick employs a regression discontinuity (RD) design in birth date to account for the aforementioned selection challenges that would undermine causal interpretation of the policy effect. She finds a clear positive effect of UPK on 4 year old preschool enrollment, but her results on maternal outcomes contrast with Gelbach's. Finding small and statistically insignificant policy effects on maternal employment, weeks worked, and hours worked per week last year (-0.05 percentage points for employment)[9], Fitzpatrick concludes that the income effect may dominate in this case, inducing mothers to reduce their participation on the intensive margin.

My study adds to this evidence in various ways. The NYC policy's recent time frame and my data's temporal variation differentiate my study from the preceding empirical analyses, which use Decennial Census data from 2000 and earlier. Specifically, studying the period since 2010 allows me to capture how a large implicit wage subsidy affects maternal labor supply in a new generation of U.S. women, one whose labor force participation and fertility decisions have diverged from those of older generations.[10] In light of these changing trends, there is little reason to expect the magnitudes or even the signs of my effects to align with those in other studies, yet results from a more recent setting are likely better predictors of responses in the future. Furthermore, since I estimate effects using American Community Survey (ACS) data from multiple survey years (2010-2016), my results are less likely to reflect idiosyncrasies of a particular year of data, which contrasts with the aforementioned studies. Finally, unlike the research designs used by Gelbach (2002) and Fitzpatrick (2010), my DID design does not identify the policy effect from a sample of mothers whose children were born at a certain time in the year. The limitation of that identifying variation is that the mothers in the sample may not generalize to the broader population of mothers in the institutional setting (whose children were born at other times in the year). Since identification in the DID design does not rely on this birth date variation, my estimates may better represent the behavior of the average mother in the target population.

The remainder of the paper is organized as follows. In Section 2 I discuss the policy

[9]Fitzpatrick's confidence intervals suggest a range of maternal employment effects from -3 to 4 percentage points.

[10]The labor force participation rate of U.S. women peaked in 2000 after having risen steadily over the second half of the twentieth century. Since 2000 it has followed a generally downward trajectory. This rate did begin to trend upward again starting in September 2015, although the current participation rate of 57.5 percent coincides with levels achieved in 1991 and is well below the all-time peak participation rate of 60.3 percent. To the extent that policymakers view the reduction in women's labor supply as stemming from families' constraints or concerns regarding preschool and early childhood care, they may be inclined to consider preschool and related childcare policies as possible levers to alter women's labor supply.

setting, followed by the empirical approach in Section 3. Section 4 contains a discussion of the data, the sample construction, and descriptive evidence. I present the results in Section 5 and discuss robustness analyses in Section 6. Section 7 contains concluding thoughts.

2 Institutional Setting

2.1 The New York City Preschool Market

New York City residents interested in sending their 4 year old child to formal preschool (henceforth referred to as pre-K)[11] may choose from publicly funded or privately funded options. The publicly-funded pre-K alternatives fall within three broad categories: (1) traditional New York City public schools, such as district schools or zoned schools; (2) charter schools, which are independent public schools that operate under a contract (or charter) of up to five years[12]; and (3) New York City Early Education Centers, which are private firms that contract with the NYC Department of Education (DOE) to administer a Pre-K For All program (additional details provided below).

2.2 Pre-K For All: Policy Overview

When Bill DeBlasio ran for New York City mayor in 2013, implementing universal Pre-K was his major campaign promise.[13] In August 2013, just before the school year began, few New Yorkers could have predicted that the universal pre-K idea would become a reality for their City (or, for that matter, that it would be available the following school year), as DeBlasio had not even won the primary yet. The rapid speed at which the City was able to deliver on DeBlasio's promise was striking: The DOE drastically increased the number of full-day pre-K seats in 10 short months.[14]

[11]Generally, "preschool" is the term used to refer to any formal schooling for children under the kindergarten-eligible age, while "pre-Kindergarten" or "pre-K" refers specifically to preschool in the year before a child will attend kindergarten and which is designed to get the child "kindergarten ready." The American Community Survey, which is the primary data source for this project, uses the terms "preschool" and "nursery school" interchangeably, and considers children under age 3 to be ineligible for any type of school.

[12]Charter schools are New York City public schools founded by not-for-profit Boards of Trustees. Charter schools manage their application processes separately (although some use the NYC Charter School Center's Common Application), but they may not exclude students on the basis of disability, race, creed, gender, national origin, religion, ancestry, intellectual ability, measures of achievement or aptitude, or athletic ability. The deadline to apply for a charter school is no earlier than April 1; if the number of applicants to a charter school is more than the number of available seats, the charter school will use a random selection process (e.g. lottery) to assign students.

[13]New Yorkers likely had become accustomed to their politicians' undelivered promises about UPK. Details about prior efforts to make pre-K "universal" in New York State can be found at https://www.nytimes.com/2008/08/23/education/23prek.html.

[14]DeBlasio was elected mayor in November 2013, yet funding for Pre-K For All was not secured until March 2014. Given the short time frame for the large-scale implementation, it was necessary

The NYC DOE launched Pre-K For All (UPK) in academic year 2014-2015. The objective of the policy is to provide all 4 year old residents a full-(school) day (6 hours, 20 min)[15], high-quality preschool experience that prepares them for success in kindergarten. The expansion was implemented over a two-year period. To increase the full-day publicly funded preschool seat capacity from around 20,000 in 2013 to the target of 70,000 (in 2016), the City collaborated with the private sector to expand preschool programs already in existence, create new programs, and convert private preschool seats or entire programs to UPK seats.

As mentioned, any private firm that contracts with the DOE to provide a UPK seat is known as a New York City Early Education Center (NYCEEC). Over 60 percent of UPK providers are NYCEECs. As contractors, NYCEECs receive per-pupil funding directly from the NYC DOE in exchange for enrolling pre-K students. Thus, any student that attends a UPK-participating preschool program—whether it be publicly provided by the DOE or privately administered by an NYCEEC—receives an in-kind transfer that covers the total (tuition) cost of attendance; the primary cost of school attendance borne by families is transportation costs, and any related childcare costs for before- or after-school care.

2.3 Policy Details

The Pre-K For All program is universal: All students meeting the DOE's eligibility criteria are guaranteed a seat if they apply. Pre-K For All eligibility is determined by the child's residency and age: Any NYC resident child who is 4 years old on Dec. 31 of a given calendar year[16] can attend pre-K beginning in September of that year. This means that when UPK began in September 2014, all NYC children born in 2010 received the UPK offer to enroll (were "treated") for school year 2014-2015. Although any eligible child has guaranteed access, UPK participation is optional.

The application process for UPK is centralized, and students are assigned to Pre-K programs via a student-proposing Deferred Acceptance (DA) mechanism. Parents face a large number (over 1800!) of differentiated preschool options over which they may express their preferences by ranking up to 12 programs. Programs characteristics vary, but all must meet regulatory standards set by the NYC Department of Education and use a NY State-approved curriculum. Aside from public versus private control, some distinguishing program features include magnet schools, flexible schedules, single-sex classrooms, dual language programs, early drop-off, and after-school/extended care (for an additional charge[17]), among others.

The DA mechanism uses parents' listed preferences and the student's priority ranking to assign students to programs. A student's priority ranking is specific to each program. Priority groups and their ordering vary by program type, but there are anywhere from 5-7 priority groups for each program. In general, NYCEECs prioritize students who attended as 3 year olds, while public programs prioritize students who will have a sibling at the school

for the NYC DOE to conduct thorough outreach to parents in the spring and summer of 2014 to promote awareness of the program to eligible families (Crawford et al. (2015)).

[15] Pre-K For All programs are full-day only.

[16] These children are henceforth referred to as "4 year olds" unless otherwise noted.

[17] These additional services may also be subsidized for families meeting certain eligibility requirements.

and who live within the school's zone or district. A random lottery is used to break ties (rank students) within priority groups.

Although each eligible child who applies to Pre-K For All is guaranteed a seat, the seat offered may not be the most preferred, and it may not even be on the child's application list. The DA mechanism handles excess demand for programs using priorities, but if no seats remain at any of the programs on the child's application list, then the child is assigned to the closest program (to her home) *with available seats*. Furthermore, learning about preschools may involve a large time cost, involving research and program visits.

Given this potentially costly search process and the somewhat black-box nature of the DA assignment process, some families may choose not to participate in UPK. A family's decision may be between a certain outside option—e.g. securing early (and paying for) a private preschool seat that meets their quality standards—and a potentially costly search process and uncertain UPK preschool assignment that is also tuition free. Risk-averse families who have the resources for private preschool may forego the UPK application process altogether. Families who are less risk-averse or who face lower search costs may initially intend to enroll their child in UPK if their expected payoff from participating is higher than that of the outside options (e.g. private preschool, charter schools, home-based schooling, or no formal preschool). Yet in the end, a family may choose the outside option if the offered seat does not yield high enough value.

3 Empirical Approach

I employ a difference-in-differences research design to estimate the causal effect of NYC's universal pre-K introduction on mothers' labor force participation and usual hours worked. My institutional setting lends itself to a DID design. As detailed in the beginning of Section 2.3, mothers are treated (offered UPK) only if their child is a NYC resident and is 4 years old. These residency and age criteria induce multiple samples of ineligible mothers that can be used as comparison groups to estimate the policy effect.[18]

To motivate the DID design, consider the effect that would be detected using a model

[18] As with many policies that use a strict rule to determine treatment assignment, UPK's December 31 birthday cutoff creates exogenous variation in age eligibility for students with birthdays around the cutoff. Although such variation is often suggestive of a regression discontinuity (RD) design in birth date (as is pursued in Fitzpatrick (2010)), this empirical approach is not feasible in my setting due to data limitations. The RD design would require (1) a considerably larger NYC sample than the ACS offers, and (2) that I observe the child's exact date of birth, a variable that is not available in the public-use ACS data. Even so, estimation of the policy effect with a DID design confers its own advantage relative to an RD design. In particular, the DID may better capture the effect of the target population. Since the effect identified by an RD is local to mothers whose children have birthdays around the cutoff, these mothers have the oldest and youngest children in a cohort, and their participation decisions may differ relative to mothers of children born throughout the year. (For instance, families who aim to have children in time to claim child tax credits by the end of the tax year may differ from families who have children at other times during the year.) Since identification in the DID does not rely on this narrowly defined subset of mothers, it may better capture the average effect of the policy on mothers of 4 year olds.

with only a single difference. Such a model could be specified using one of many potential contrasts, leveraging variation within a cohort or across cohorts to identify the treatment effect. For example, one could model the effect of UPK using variation in NYC mothers' 2014 labor force participation across age groups, comparing mothers of 4 year olds with mothers of 3 year olds. The shortcoming of this design is that mothers' participation decisions presumably correlate with the child's age. As a result, the effect identified by the post period difference in mean outcomes between treated (child's age = 4) and comparison mothers would include the true policy effect and a bias term representing the baseline (non-treated state) difference in outcomes between mothers of 4 year olds (eligible) and mothers of 3 year olds (ineligible/comparison). The DID overcomes this single-difference design limitation by leveraging data in the pre-policy period to remove the aforementioned bias component from the estimated policy effect.

3.1 Difference-in-Differences Research Design

In an ideal setting, I would observe a group of mothers that is identical to the treatment group on all dimensions except that they are not treated; in other words, the treatment would be randomly assigned among mothers of 4 year olds in New York City. The nature of any universal policy, however, is that the entire population is offered treatment, so in such settings the ideal comparison group never exists. Intuitively, one can think of the DID design as attempting to mirror identification in a randomized controlled trial by constructing an appropriate comparison group against which to measure the effects of the policy on the treated sub-population.

Accordingly, identification of the policy effect in the DID design depends importantly on the choice of comparison group. The comparison group need not have the same levels in potential outcomes as the treated group. Rather, the appropriateness of a comparison group stems from whether the second contrast (or difference) reliably removes the first contrast's underlying bias in potential outcomes, which will occur when the level of bias is the same for each of the differences. When contrasts involve changes over time, this "bias stability" assumption is often referred to as the "parallel" or "common" trends assumption, and implies that absent the policy, the outcomes between the eligible and comparison groups would have followed the same trends over time.

Accordingly, my objective for identification is to construct a comparison group that is (i) not affected by the UPK offer and (ii) whose expected realized outcomes in the post period would equal the expected counterfactual (unobserved) outcomes of the eligible population had they not received treatment in the post period. In my empirical setting, mothers whose child just fails eligibility on one of the required dimensions (age, residency) cannot receive the treatment (satisfying criteria (i) above), yet may be reasonably comparable to the group of treated mothers (criteria (ii)).

To make the comparison mothers as similar as possible to the treated mothers, I define my comparison groups as (1) mothers of 4 year olds who reside outside New York City's geographic boundary but within the NYC metropolitan area (MSA comparison) and (2) NYC mothers of 7 year olds (NYC/age comparison). Each of these has strengths and weaknesses as a counterfactual group, which I discuss in Sections 3.1.1 and 3.1.2 below.

To preview my analytical approach, I estimate a separate DID model using each of the different comparison samples. Each model uses time as one of the differences; what varies across models is the comparison sample and, hence, the implicit meaning of ineligibility.

3.1.1 Comparison Group 1: Residency Contrast

My first comparison group comprises mothers with 4 year olds who live in the New Jersey counties that neighbor New York City ("MSA mothers"). Their children satisfy the UPK age requirement but are ineligible for the UPK offer because they reside outside the NYC city limits. Empirically, mothers' labor market patterns change with their children's ages. Since mothers with a 4 year old presumably share similar parental obligations, schedules, consumption patterns, and constraints induced by that child's age-specific needs, I believe the parallel trends assumption is more likely to hold if the comparison group mothers also have a 4 year old.

Among non-NYC resident mothers of 4 year olds, those who also live in close geographic proximity to the eligible mothers—e.g. just outside NYC proper but within the MSA—may constitute an effective comparison for the eligible mothers.[19] Since they reside in the same local labor market as the eligible mothers, these comparison moms should experience the same local economic shocks. Moreover, the common experience of raising a 4 year old should elicit a response to these shocks that is similar to that of the eligible mothers.

3.1.2 Comparison Group 2: Age Contrast

For the second comparison group, I begin by considering mothers who reside in NYC but whose child is either slightly too old (5, 6, or 7 years old) or slightly too young (2 or 3 years old) to qualify for the UPK offer. Just as the common experience of parenting a 4 year old may confer advantages for the residency contrast described above, there also may be an advantage to limiting the comparison sample to mothers who reside in NYC. In particular, using only NYC resident mothers for comparison should reduce selection-based differences in labor market proclivities between eligible and comparison mothers that arise from the decision to live in NYC proper. If the factors that induce geographic selection also lead to residency-based differences in maternal labor supply responses to temporal shocks, then the parallel trends assumption may be violated for the previously discussed MSA comparison group. Restricting the sample to NYC residents, then, may provide a robustness to jurisdiction-based differences in both the public policy landscape and the response to economic fluctuations.

The differences in ages, however, may be a weakness for this comparison group. Although mothers with 2, 3, and 4 year olds all have children in the pre-elementary age range, the age differences might correspond to important differences in how mothers trade off labor and leisure, as well as differences in the relative prices of home versus market provided childcare. In a similar vein, mothers of 5, 6, and 7 year olds likely participate in the labor market

[19] One may be concerned that mothers who choose to live within the five boroughs are markedly different than mothers who live outside the boundaries, especially if the cost of living differences are large. The key assumption here is that even if these mothers appear different based on observable demographics, as long as their differences remain constant over time (i.e., they share and respond equivalently to economic fluctuations and unobserved shocks), they can be used to remove the non-policy time effects from the treated group.

at higher rates than mothers of 4 year olds (absent the UPK offer) since the K-12 public schooling system offers an implicit childcare subsidy for at least 6 hours and 20 minutes a day during the academic year. Although baseline (pre period) differences in participation rates between eligible and comparison mothers is not itself a problem for the DID, if time-varying factors affect labor force attachment differently for mothers of 2-3 and 5-7 year olds than for mothers of 4 year olds, then bias stability could fail.

A further concern with using mothers of 2-3 year olds for comparison is that these mothers will be eligible for the UPK offer in the near future. The anticipation of the policy might lead these mothers to re-optimize, so that their labor market behaviors are affected by the UPK introduction even before their children are old enough to be eligible to participate. To avoid the complications that such anticipation would cause for the analysis, I consider only mothers of slightly older children for my comparison group. Finally, due to measurement error arising from data limitations (discussed further in Section 4.3 below), I exclude mothers of 5 and 6 year olds and use only NYC mothers of 7 year olds for my age comparison group.[20]

Table 1 summarizes the eligible and comparison samples based on the criteria for UPK eligibility.

Table 1: Construction of Comparison Group Mothers Using Eligibility Criteria

	Residency	
Child's Age	NYC	MSA, Non-NYC
4 yrs old	*UPK Eligible*	MSA Comparison (DID 1)
7 yrs old	Age Comparison (DID 2)	Added to DID 1 and 2 for DDD

3.2 Econometric Models

My econometric models follow the standard difference-in-differences regression specification in the cross-sectional setting, which includes an indicator for treatment eligibility, an indicator for the policy period, their interaction, and a set of time dummies. My main specification also includes individual covariates, as I believe the parallel trends assumption holds conditional on covariates, but not necessarily unconditionally, because of observable demographic differences between the treatment and comparison groups.

Let mothers be subscripted by i, with $i = 1, ..., N$, while $t_i \in \{2008, \ldots, 2016\}$ denotes the academic year corresponding to i's interview date, t_i^I, with $t_i^I \in \{$Jan. 1, 2010, ...,

[20] A third comparison group that I considered was NYC mothers of 7 and 9 year olds. The reason for adding mothers of 9 year olds is that this cohort of children never would have been treated (offered UPK) in any of the post period years. However, since the bias stability assumption does not appear to hold in the pre period for this group, I concluded that the DID design with this comparison group likely is not valid. Results for this comparison group are available by request.

Dec. 31, 2016}.[21] Each individual is observed only once. Let Y_i denote i's outcome variable, either an indicator of labor force participation ($Y_i \in \{0,1\}$), or the ("continuous") usual hours worked per week ($Y_i \in [0,70]$). Furthermore, let

- $Post_i \in \{0,1\}$ indicate whether i is interviewed in the post-policy period (once UPK has begun), equal to 1 if $t_i \geq 2014$, 0 otherwise;
- $Age_i \in \{4,7\}$ denote the age of i's child on Dec. 31 of academic year t_i;
- $Four_i \in \{0,1\}$ be an indicator of whether i's child is 4 years old on Dec. 31 of t_i, that is,
$$Four_i = \begin{cases} 1 & \text{if } Age_i = 4 \\ 0 & \text{otherwise} \end{cases}; \text{ and}$$
- $NYC_i \in \{0,1\}$ denote whether i is a resident of NYC (=1) in academic year t_i.

Since a child must satisfy both the age and residency criteria to be eligible for Pre-K For All, the eligibility variable is defined as

$$Elig_i = \begin{cases} 1 & \text{if } Four_i(t_i) = 1 \text{ and } NYC_i(t_i) = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Note that an individual can satisfy these eligibility criteria whether she is observed in the pre or post period. The remaining variable of interest is the "treatment offer" indicator, which is constructed by interacting the eligibility and post period indicators:

$$T_i = Elig_i \times Post_i.$$

This variable is equal to 1 if i's child was eligible for UPK in the post period, and 0 otherwise. Finally, let X_i denote covariates for mother i in period t.[22]

My difference-in-differences model is given by the following regression equation:

$$Y_i = \delta + \alpha T_i + \theta Elig_i + \gamma Post_i + \sum_{s \in S} \phi_s 1_{\{t_i = s\}} + \tau X_i + u_i, \tag{1}$$

where u_i is assumed to be uncorrelated with the regressors.[23]

[21] To be clear, if $t_i^C(t_i^I) = year(t_i^I)$ denotes the calendar year in which i's household is interviewed, then the academic year t_i is given by

$$t_i = \begin{cases} t_i^C - 1 & \text{if month}(t_i^I) \in \{\text{Jan., } \ldots \text{, Aug.}\}, \\ t_i^C & \text{if month}(t_i^I) \in \{\text{Sept., } \ldots \text{, Dec.}\}. \end{cases}$$

That is, the academic year begins September 1 of t_i^C and ends on August 31 of $t_i^C + 1$.

[22] These variables include mother's educational attainment, marital status, race, Hispanic origin, age and age^2, number of own children in the household, age of youngest own child in the household, disability status, a working age indicator, and the child's birth quarter.

[23] S is the set of sample years with one year omitted in the pre and post period.

Equation (1) is a Linear Probability Model for the binary participation outcome and a standard linear regression model when the outcome is the mother's usual hours worked per week. The coefficient on T_i is the parameter of interest: α is an intent-to-treat (ITT) estimand capturing the effect of the UPK offer on the labor force participation or hours worked of treated mothers. Recall that the key assumption for identification of α is that the unobserved time effects are the same for the eligible and comparison mothers ("parallel trends" or "bias stability"). The eligible and comparison groups may come from different underlying populations, but the time-variant factors must affect them in the same way.

A clarification should be made regarding the selection of the sample. While the set of individuals satisfying $Elig_i = 1$ is the same across models, the rest of the observations included in the sample will depend on the contrast. Accordingly, the meaning of $Elig_i = 0$ changes depending on which comparison group sample is used. When estimating Equation (1) using the MSA comparison group, the set of ineligible ($Elig_i = 0$) individuals is mothers of 4 year olds who live in the MSA but not in NYC (residency contrast), whereas when estimating Equation (1) using the NYC comparison group, the set of ineligible individuals comprises NYC mothers of 7 year olds (age contrast). To make these separate DID models explicit, the general regression given by Equation (1) can be rewritten as two estimating equations, each with a corresponding sample:

DID 1: Residency Contrast (Comparison Sample: MSA Mothers of 4 Year Olds)

$$Y_i = \delta + \alpha T_i + \theta NYC_i + \gamma Post_i + \sum_{s \in S} \phi_s 1_{\{t_i = s\}} + \tau X_i + u_i, \tag{2}$$

where $T_i = Elig_i \times Post_i = NYC_i \times Post_i$ and the sample is restricted to mothers for whom $Four_i = 1$.

DID 2: Age Contrast (Comparison Sample: NYC Mothers of 7 Year Olds)

$$Y_i = \delta + \alpha T_i + \theta Four_i + \gamma Post_i + \sum_{s \in S} \phi_s 1_{\{t_i = s\}} + \tau X_i + u_i, \tag{3}$$

where $T_i = Elig_i \times Post_i = Four_i \times Post_i$ and the sample includes only mothers for whom $NYC_i = 1$.

3.3 Estimation

I estimate the regressions in Equations (2) and (3) using OLS. As mentioned above, I believe the parallel trends assumption holds conditional on covariates, so the main specification of interest is the one that includes covariates; nevertheless, for comparison I also present estimates of the models that omit the covariates X_i.[24]

[24]If covariates are unnecessary for identification, then their role should just be to reduce residual variation, resulting in smaller standard errors. If the DID model is valid only after conditioning on covariates, then the estimated coefficients would vary between the two specifications.

Since there are multiple years in the post-policy period, I conduct separate analyses for two time frames. The first uses data from 2010-2014 and thus estimates the effect of the policy in the first year only. The second time frame adds the additional post-period years (2015-2016) to estimate the policy effect, and accordingly uses data for 2010-2016. The reason for considering both time frames is that the policy took two years to implement fully. While the number of children participating in full-day public pre-K increased in the first year, this came at the cost of a reduction in the number of half-day pre-K seats available, as the expansion relied heavily on converting half-day seats to full-day seats. Accordingly, the first year may have seen an overall reduction in the number of children that could be served by pre-K, even if more children could attend full-day pre-K. This change in the availability of seats may have induced a different effect on mothers' participation than would occur once the policy was fully implemented.

For each model, I take two different approaches for the estimation of standard errors. As is well known, the errors in the Linear Probability Model (LPM) are heteroscedastic due to the binary nature of the outcome, so at a minimum any LPM model should be estimated using Huber-White heteroscedasticity robust standard errors. However, in DID settings a standard concern is serial correlation in the errors. I also am concerned about spatial correlation of errors among individuals who live in the same neighborhood, especially since employment and pre-K options may vary substantially across neighborhoods. To address these two concerns, I estimate the results with standard errors clustered by the Public Use Micro Area (PUMA) in which the household is located.[25] By clustering at the level of the PUMA, I also allow the errors to be correlated over time for individuals living in the same PUMA. I believe that this clustering approach is preferred to the more restrictive error structure that only accounts for individual heteroscedasticity, so results in the body of the paper use standard errors clustered by PUMA.[26]

3.4 Triple Differences Design (DDD)

As previously discussed, each of the aforementioned DID comparison groups offers advantages in estimating the policy effect. An alternative model specification could leverage the complementary strengths offered by each contrast to estimate a single model that includes differences over time across age and residency. This Difference-in-Difference-in-Differences or "triple differences" (DDD) design uses a sample consisting of all MSA and NYC mothers of 4 and 7 year olds and adds an additional set of interactions between the two eligibility variables, $Four_i$ and NYC_i, and the policy indicator $Post_i$. The DDD model is specified as:

DDD: Age and Residency Contrasts (Comparison Sample: NYC Mothers of 7 Year Olds and MSA Mothers of 4 and 7 Year Olds)

[25]PUMAs are a geographic unit defined by the U.S. Census Bureau. New York City contains 55 PUMAs (10 in Bronx, 18 in Brooklyn, 10 in Manhattan, 14 in Queens and 3 in Staten Island). Areas exceeding 200,000 residents are divided into as many PUMAs of 100,000+ residents as possible. (See the IPUMS documentation at https://usa.ipums.org/usa-action/variables/PUMA#description_section.)

[26]Nevertheless, the results are qualitatively unchanged when I use robust standard errors.

$$
\begin{aligned}
Y_i = \ \delta \ + \ & \alpha(NYC_i \cdot Post_t \cdot Four_i) \\
& + \beta_1(NYC_i \cdot Post_i) + \beta_2(NYC_i \cdot Four_i) + \beta_3(Post_i \cdot Four_i) \\
& + \theta_1 NYC_i + \theta_2 Four_i + \gamma Post_i + \sum_{s \in S} \phi_s 1_{\{t_i = s\}} + \tau X_i + u_i, \qquad (4)
\end{aligned}
$$

where, as before, α is the parameter of interest and u_i is uncorrelated with the regressors.

The contribution of the DDD is to remove time-varying effects shared by mothers of a given age group, as well as time-varying effects shared by mothers in a given geographic location. Accordingly, the DDD estimator of the policy effect is consistent if there are no unobservable determinants of maternal labor supply outcomes that coincide with the introduction of the UPK policy and that differentially affect mothers of either age group in NYC or the surrounding MSA. In other words, identification requires that any time-varying, location-specific unobservables that affect the labor supply decisions of mothers with 4 year olds also are shared by mothers of 7 year olds, and that any time-varying, age-specific unobservables affecting the labor supply decisions of NYC mothers also affect those of MSA mothers.

4 Data and Descriptive Evidence

4.1 Data Source and Variables

I use data from the American Community Survey (ACS), one of the U.S. Census Bureau's ongoing household surveys. The ACS collects rich household- and individual-level data on a broad range of subjects—including demographics, employment, income, education, migration, and disability—which previously was obtained from the (now discontinued) Long Form of the Decennial Census.[27] The ACS is administered daily through internet, paper, phone, and in-person surveys, and its information is reported on an annual (calendar year) basis. The sampling frame includes about 3.54 million addresses each year and covers 91.6 percent of the U.S. population (2016).[28] I use 1-year samples of the Public Use Microdata Sample (PUMS) from calendar years 2010 through 2016.

The outcome variables of interest are the respondent's current labor force participation status and usual hours worked per week. The relevant demographic variables include each mother's age, race, Hispanic origin, and educational attainment (highest grade completed).[29]

[27]In this setting, the benefit of using the ACS to study maternal labor market responses, rather than an alternative labor survey like the CPS, is that the ACS is a representative survey even at smaller geographic levels (and offers such data), which would not be the case for the CPS.

[28]See https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/sample-size/ and https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/coverage-rates/.

[29]For the sample summary statistics I also include mother's current educational enrollment and an indicator that the respondent received public assistance income, but these variables are not included in the regressions.

I also account for two factors that should affect the mother's degree of labor market attachment: Whether she is of working age (20-60), and whether she reported having a disability that causes her to experience independent living difficulty. Due to my limited sample size, instead of restricting the sample to include only working age mothers without a disability, I opt to include these as covariates in the analysis. Finally, geographic variables used either directly in the regressions or to construct the analysis sample include current and prior year state and county of residence, a metropolitan status indicator, and a finer identifier of geographic location approximated by the public use micro area (PUMA) (e.g. Upper West Side, Manhattan).

For children, the relevant demographic and preschool related variables available in the ACS include age, quarter of birth, school attendance status, the type of school the child is attending (public or private/home schooled), the school grade if currently attending (or attended within the prior three months)[30], and educational attainment (highest grade completed). The only child variables used for the analyses are age, quarter of birth, school attendance, and grade attended.[31]

4.2 Construction of Samples

To construct my analysis sample, I begin by obtaining the individual-level ACS records for all respondents age 0-18 who reside in the New York City Metropolitan Statistical Area (MSA).[32] For each of these records, I attach the corresponding record of the child's mother, which is included if she lives in the same household.[33]

As discussed in Section 3.2, since I estimate multiple DID models using age and geographic contrasts, I define a separate sample for each model specification. For the first sample, I include only mothers who are residents of the New York City MSA and have a 4 year old.[34] The second sample contains only mothers who are NYC residents, but whose

[30]This wording of the question in the ACS should imply that most elementary and secondary school-aged children in school during the academic year respond as attending school, even if they are interviewed in the summer. Indeed, in the ACS data for NYC, 98.12 percent of children aged 6-15 are reported as attending school.

[31]I have examined the effect of UPK on enrollment in preschool since it is the mechanism through which I expect to detect an effect on maternal labor supply outcomes. (Appendix B contains figures of NYC preschool enrollment over time for children age 3-5.) I find clear evidence of an increase in preschool enrollments after the policy, which is consistent with administrative enrollment numbers from the NYC DOE. Nevertheless, since the ACS does not ask about full-time or part-time enrollment status, nor about whether the child is a UPK participant, the ACS data offer little toward understanding how preschool enrollment behaviors changed.

[32]All ACS PUMS data is accessed via the IPUMS-USA database hosted at the University of Minnesota (See Ruggles et al. (2019)).

[33]Children are dropped from the final sample if no mother is present in the household/the mother's information is missing. For children who have more than one mother (i.e., live in a same-sex household), I use only the information from the first mother listed. This is a very small percentage (0.35 percent) of cases: for 2010-2016, only 36 of the 10,230 NYC 4 year olds containing the first mother's information have a second mother with age information.

[34]The comparison mothers are restricted to be New Jersey residents in MSA counties surrounding

child is 4 years old or 7 years old. The third sample is used for the DDD, combining the mothers from the first two samples and also adding MSA mothers with 7 year olds.[35] For all samples, I drop duplicate observations of mothers resulting from their having more than one child of the same age (e.g. twins or triplets). For the samples that use mothers of 7 year old children as a comparison group, duplicates are also a concern, as it is possible that a mother appears in both the eligible and comparison groups. I address this problem by keeping the maternal record for the youngest child, which implies that the mother is assigned to the eligible group.[36]

Further restrictions could be made to each sample based on the age of the mother's youngest child, regardless of whether that child is 4 or 7 years old. Specifically, I could discard records of mothers who have any children younger than age 4. At first glance, one might expect greater salience of the UPK offer for the labor market decisions of mothers who do not still need to arrange childcare for another younger child when their 4 year old attends free pre-K. For instance, a nonworking mother with a 4 year old and a 1 year old may be less likely to change her labor force status in response to UPK, even if her 4 year old attends, since she will still have to provide childcare for an infant. On the other hand, the reduction in childcare needs from two children to one child may reduce the relative price of market-based childcare enough to induce the mother to reenter the labor market. It is unclear which of these effects would dominate. In any case, it seems too restrictive to reduce the sample only to mothers whose youngest child is 4 or 7, and more appropriate to include all eligible mothers. An additional reason not to exclude eligible mothers is that it would reduce my (already modest) sample size and, thus, my statistical power.

4.3 Data Challenges and Solutions

A relatively innocuous limitation of my data is that the ACS dataset does not have a variable indicating UPK participation, only whether the student is enrolled in "public school" or "private school/home school," and the corresponding grade.[37] With perfect data, I would know detailed information about each eligible child's preschool enrollment decision: When he begins any formal schooling, if and when he begins participation in UPK, the type of school he attends (public, charter, NYCEEC, other non-UPK private), tuition paid, and whether he attends half-time or full-time. Yet because UPK-participating schools may be private, and because charter schools are public but are not UPK participants, with only ACS data I cannot distinguish which children are participating in the UPK program versus

NYC, since mothers in Long Island and Westchester are likely to be less comparable, a hunch that I confirm by examining the pre period trends when they are included in the MSA comparison group.

[35]Since the set of eligible mothers is unchanging across the different model specifications, in any given sample, the remaining "ineligible" mothers constitute the comparison group for that sample.

[36]An alternative approach could be to keep the record for the mother's oldest child in the sample. I keep the record of the youngest child in the sample because I am likely to have fewer treatment group records, and because the age of the mother's youngest child seems most likely to induce the mother's participation constraint to bind or reduce her hours worked.

[37]Nursery school and preschool are considered one category.

another preschool alternative.[38] Although this limitation precludes more detailed analysis of the relationship between UPK enrollment and mothers' labor force participation, it does not affect my ability to answer the research question.

Unfortunately, I face various other data limitations that, if unaccounted for, may induce measurement error in the age variable, academic year variable, and consequently, in the determination of whether a record should be included in the analysis sample. These errors could bias my estimates of the policy effects. I discuss each of these limitations and the corresponding errors induced, followed by my solution.

4.3.1 Unobserved Birth Date

The first data limitation I face is that the ACS withholds individuals' true birth year from the public-use dataset. The birth year fully characterizes the child's age on December 31 of a given academic year, which indicates whether he is eligible for UPK in that academic year and should be included in the sample. Although the IPUMS database contains a variable called "birth year," it is imputed using the interview year and the child's age (in years) at interview. As a result, the birth year variable is incorrect for individuals who are interviewed before their birthday in that calendar year.[39]

Given that the birth year is unobservable, I instead must use the observed age at interview as a proxy for true age on Dec. 31 of a given academic year. If this were the only data limitation (and if no adjustments are made), then eligibility assignment using observed age (e.g., selecting only 4 year olds for the MSA DID) would induce misclassification errors in the age variable. This error would propagate to the selection of ACS records to be included in the sample. The corresponding misclassification can be categorized as incorrectly excluding an individual from the sample (error of exclusion) or as incorrectly including an individual in the sample (error of inclusion). The details of these errors are as follows:

Errors of Exclusion:

1. Children interviewed at age 3 who will be 4 years old on Dec. 31 of the corresponding academic year will be misclassified (for eligibility purposes) as 3 years old and, accordingly, incorrectly excluded from the sample.
2. Children who turned 5 years old before the interview but who were 4 years old on Dec. 31 of that academic year (which would be the prior December) will be misclassified as 5 years old and incorrectly excluded from the sample.

Errors of Inclusion:

[38] Nevertheless, because the ACS provides a representative sample of New York City, the data enable me to observe overall changes in NYC residents' preschool choices among public, private, and "no participation" alternatives.

[39] As an example, suppose a child was born on November 1, 2010 and her family was interviewed sometime January 1 through October 31, 2014. Her observed age at interview is 3, so her imputed birth year variable is $2014 - 3 = 2011$, while the correct birth year is 2010.

1. Children interviewed at age 4 who will be 5 years old on Dec. 31 of the corresponding academic year will be misclassified (for eligibility purposes) as 4 years old and incorrectly included in the sample.

4.3.2 Unobserved Academic Year

The second major limitation of the data is that I cannot observe the ACS interview date. This variable is needed to know the academic year in which a household was interviewed, which, together with the child's birth date, determine whether the mother should be included in the analysis sample and for which year. Specifically, since my research design compares the participation of mothers when their child is pre-K aged, the only records of interest are those of mothers interviewed during the academic year in which their child is eligible.[40] (I believe it is easiest to think of the age eligibility assignment as being determined in two steps: First, determine the academic year in which the household is interviewed. Second, determine the child's age on the unique Dec. 31 spanned by that academic year. That gives the correct age to be used to determine whether a child is 4 or 7 years old and, thus, relevant for one of my analyses.)

Although ACS interviews occur daily, only the calendar year of interview is publicly available. The ACS has 12 monthly independent samples, so for a given calendar year, 2/3 of households will have been interviewed during the prior academic year (January 1 through August 31), while the remaining 1/3 will have been interviewed from September 1 through December 31, when the academic year and calendar year coincide. If I observed the birth year, then it would be straightforward to correct my regression estimates using the assumption that interview dates are distributed uniformly across the calendar year.

For example, consider calendar year $t^C = 2014$. Since the academic year begins September 1, let $F = 1$ denote a household interviewed in the fall (September-December), while $F = 0$ if the interview date was sometime in January-August. For any outcome Y, I can compute the mean for the calendar year, $E(Y_{t^C})$, which by the law of iterated expectations can be decomposed as

$$\begin{aligned} E(Y_{t^C}) &= E(Y_{t^C}|F=1) \cdot Pr(F=1) + E(Y_{t^C}|F=0) \cdot Pr(F=0) \\ &= E(Y_{t^C}|F=1) \cdot Pr(F=1) + E(Y_{t^C}|F=0) \cdot (1 - Pr(F=1)) \\ &= [E(Y_{t^C}|F=1) - E(Y_{t^C}|F=0)] \cdot Pr(F=1) + E(Y_{t^C}|F=0) \\ &= [E(Y_{t^C}|F=1) - E(Y_{t^C}|F=0)] \cdot \frac{1}{3} + E(Y_{t^C}|F=0), \end{aligned} \tag{5}$$

where $Pr(F = 1) = \frac{1}{3}$ by the assumption that interviews are distributed uniformly across months.

To put this in the DID framework, think of Y as a difference in outcomes between calendar years t^C and $t^C - 1$ conditional on UPK eligibility status. That is,

$$E(\Delta Y_{t^C}^{Elig}) \equiv E(Y_{t^C}|Elig=1) - E(Y_{t^C-1}|Elig=1) \tag{6}$$

[40] In other words, the analysis sample should exclude mothers who were interviewed in the academic year when their child is 3, because this is the academic year before the child could enroll in UPK, and when their child is 5, because in this academic year they should be in kindergarten.

for eligible mothers and

$$E(\Delta Y_{t^C}^{Comp}) \equiv E(Y_{t^C}|Elig = 0) - E(Y_{t^C-1}|Elig = 0) \tag{7}$$

for comparison mothers.

Then Equation (5) can be written as

$$E(\Delta Y_{t^C}^{Elig}) = [E(\Delta Y_{t^C}^{Elig}|F = 1) - E(\Delta Y_{t^C}^{Elig}|F = 0)] \cdot \frac{1}{3} + E(\Delta Y_{t^C}^{Elig}|F = 0) \tag{8}$$

for eligible mothers (and similarly for comparison mothers). Recall that UPK begins in the fall. Accordingly, the term $E(\Delta Y_{t^C}^{Elig}|F = 0)$ will be the same for eligible and comparison mothers under the parallel trends assumption, because neither group should be affected by the policy before it has been introduced. Therefore,

$$\begin{aligned} E(\Delta Y_{t^C}^{Elig} - \Delta Y_{t^C}^{Comp}) &= E(\Delta Y_{t^C}^{Elig} - \Delta Y_{t^C}^{Comp}|F = 1) \cdot \frac{1}{3} \\ \Rightarrow E(\Delta Y_{t^C}^{Elig} - \Delta Y_{t^C}^{Comp}|F = 1) &= 3 \cdot E(\Delta Y_{t^C}^{Elig} - \Delta Y_{t^C}^{Comp}). \end{aligned} \tag{9}$$

With a mismeasured academic year variable, Equation 9 says that, under the assumption of uniformly distributed ACS interview dates, the true effect from the policy will equal the estimated DID effect scaled by 3. In other words, absent missing data on birth year, the missing academic year causes the estimated policy effect to be attenuated to 1/3 of the true effect.

4.3.3 Error Induced by the Interactions of the Mismeasured Proxy Variables

Unfortunately, the missing birth year complicates the error induced by the missing interview date information. In this DID setting, using observed variables (child's age at interview and child's calendar year of interview) as proxies for the true unobserved variables can result in misclassification of the age (and age eligibility) variable, academic year variable, or both. These errors, in turn, can cause errors of inclusion, exclusion, or misclassification (in age or academic year), or any combination of them. To complicate things further, the type and extent of errors will depend on which ages are selected for the analysis sample.

Example

To fix ideas, I discuss a simple example that shows how the two sources of missing data interact with one another. Consider a child who is age 4 at interview and whose birthday for that calendar year preceded the interview date. If his household is interviewed sometime from September 1 to December 31, then the child is age eligible for that academic year and should be included in the sample. More specifically, the calendar year coincides with the academic year, and the observed age corresponds to the eligible age for the academic year. Accordingly, assignment using observed proxies would cause no error under this parameterization.

On the other hand, if instead the household is interviewed sometime between January 1 and August 31, then the child is not age eligible for the corresponding academic year because he would have been age 3 on the prior December 31. In other words, if birth year and academic year were both observable, the child would be excluded from the analysis sample. Note, importantly, that this conclusion implies that the child also is ineligible for the subsequent (incorrect) academic year (whose value coincides with the observed calendar year). Although he would be age 4 on December 31 of that academic year, his household was interviewed in the prior academic year, before he could have started pre-K, so the corresponding information about the mother's participation is irrelevant given my DID design. Yet if the (calendar) year and observed age variables are used as proxies, then this child's mother would be assigned incorrectly to the eligible sample for the following academic year. Thus, using proxies would induce an error of inclusion.

Furthermore, consider an alternative parameterization wherein the child still is observed as 4 years old, but instead he is interviewed before his birthday for that calendar year. Then, if his household was interviewed sometime from January 1 to August 31, the child would be eligible for that academic year (as he would have been 4 years old on the preceding December 31). Using proxy variables in this case will lead to the child being included (correctly) in the analysis sample, but with a mismeasured academic year variable. Alternatively, if the household was interviewed sometime September 1 through December 30, then the child would be 5 years old on December 31 of that academic year, and thus should be excluded from the sample. However, using proxy variables will lead to incorrect inclusion of this record in the analysis.

4.3.4 Solution

Since the missing variables affect who is included in the analytical sample, my problem differs from the standard problem discussed in the literature on misclassification errors in treatment effects settings (e.g., Battistin and Sianesi (2011) or Lewbel (2007)). These papers consider treatment assignment errors in conditional expectations regressions that satisfy the unconfoundedness assumption, which is a different problem than a record being included in the sample and consequently assigned to an eligible or comparison group when it should be in neither (or, alternatively, being excluded from the sample and missing from either the eligible or comparison group). The variables measured with error interact with the construction of each of the 4 groups[41] that make up the DID design, so the missing variables induce an additional level of complication for the DID framework.

I consider various adjustments to correct for the error that would be created by the missing variables. These adjustments make use of two pieces of auxiliary information in the data: The grade in school for children currently attending (or who attended within the last 3 months) and the child's quarter of birth. I present my preferred (and implemented) approach here and discuss other possible adjustments in Appendix A.

My preferred approach, which I implement for all analyses, is designed to reduce the number of individuals incorrectly excluded from the analysis sample while also leveraging auxiliary information about birth quarter and school enrollment to remove children who are likely miscategorized. I begin by constructing the set of "4" year olds to include all observed 3 year old children born in the 4th quarter, and all observed 4 and 5 year olds. In contrast to an approach that defines the set of "4" year olds using only observed 4 year old children, this approach casts a wider net by starting with the set of all possibly eligible children.[42] I

[41]That is, (1) eligible in pre period, (2) eligible in post period (treated), (3) comparison in pre period, and (4) comparison in post period.

[42]Technically, a small subset of 3 year olds born in the third quarter also should be in the

then remove all children who are currently enrolled in kindergarten.

This approach has various benefits.[43] First, it will reduce the errors of exclusion arising from missing eligible children who had their birthday after the interview (observed 5 year olds) or before the interview (observed 3 year olds). Second, it will remove observed 4 year olds who would be included in the sample incorrectly due to error in the age variable, which could arise if their household is interviewed in the fall just before their 5th birthday. The information used to remove these children—kindergarten attendance—also is used to remove older 5 year olds that should be excluded.

The drawback of this approach is that it will not correct fully for the measurement error. By adding 3 year olds, I may include inadvertently some observations that should be excluded. Furthermore, my sample will (still) incorrectly include some 4 year olds whose households are interviewed in the academic year preceding their pre-K eligibility. Nevertheless, I believe these adjustments are an improvement upon the approach that defines eligibility using only observed 4 year olds. Moreover, in future drafts of this paper I will obtain the restricted access ACS data through the U.S. Census Bureau's Wisconsin Research Data Center (RDC), which will eliminate the measurement error challenges described here.

4.4 Summary Statistics

In this section, I briefly review the descriptive evidence on the estimation samples. Table 2 provides unconditional outcome means for each sample and period. Table 3 offers a comparison of pre period summary statistics between eligible mothers and the various comparison groups of mothers. Table 4 below and Tables 13 and 14 in Appendix B.1 show summary statistics for the eligible, MSA comparison, and NYC age comparison mothers, respectively, before and after the introduction of UPK. These covariate tables also include the differences in means (pre - post) and the corresponding t-test.

Table 2 shows that in the pre period, 63 percent of 5,984 eligible mothers participated in the labor market, and on average, eligible mothers worked 21.5 hours per week. Over the same period, the MSA comparison sample contains records for 6,012 mothers of 4 year olds in the neighboring New Jersey counties, who participate at a rate of 67 percent and work on average 23.2 hours per week. The other DID comparison group contains 1,454 NYC mothers of 7 year olds, who have greater labor supply than the mothers of 4 year olds in both groups, as might be expected since their children are of compulsory schooling age. These NYC mothers participate at a rate of 72 percent and work on average 25 hours per week. Finally, the DDD comparison mothers ($N = 9,033$) have a 69 percent participation rate in the pre period and work an average of 23.8 hours per week.

In the post period, participation rates remain the same for all comparison groups, and mothers in the MSA and DDD group increase their hours worked by approximately 1 hour per week. Hours are approximately flat for the NYC comparison mothers from pre to post

sample. But for a 3 year old born in the 3rd quarter of a given year, the probability that he should be included in the sample is only about 1 percent.

[43]Indeed, it appears that this adjustment does reduce the bias resulting from misclassification error and the errors of inclusion and exclusion in the analysis sample, as the results are a slightly larger magnitude when only observed 4 year olds are selected.

period. For the eligible mothers, however, participation rates increase from 63 to 66 percent from the pre to post period, and they work almost 2 additional hours per week, an increase from 21.5 to 23.2 hours. The relative stability in outcomes for the comparison groups is reassuring, while the jump in both participation and hours for eligible mothers provides correlational evidence that NYC mothers of 4 year olds increase their labor supply when the pre-K subsidy is made universal.

Table 3 shows that on average in the pre period, the sample of eligible mothers is approximately 34 years old, 43 percent white, 31 percent Hispanic, and has 2.5 children. About a third of women have a college degree or higher, while 24 percent attained a GED or high school diploma but no college degree. About 29 percent are single, 6 percent receive public assistance[44], and 9 percent are in school.

Moving on to the demographic composition of the comparison samples, inspection of Table 3 demonstrates that, perhaps not surprisingly,NYC residents differ in various dimensions from their "suburban" counterparts. In column (2), MSA mothers are about 1 year older, less likely to be single (19 percent of sample), more white (66 percent), less Hispanic (22 percent), less likely to be enrolled in school or receive public assistance (6 and 3 percent, respectively), and more likely to possess a college degree (51 percent). These differences are not surprising given the urban-suburban differences nationwide. Although the presence of differences in underlying demographics does not undermine the legitimacy of the DID design, it does suggest that it may be important to condition on covariates in the regressions to account for these population differences.

In contrast to the above discussion, the comparison group comprising NYC mothers of 7 year olds is fairly similar to the sample of eligible mothers. They differ in age by about 4 years (to be expected given the difference in their children's ages), and comparison mothers are slightly less white (37 percent) and slightly less likely to be in school or receive public assistance (8 and 6 percent). Additionally, NYC comparison mothers have slightly fewer children on average (2.1). This likely is due to the fact that I exclude mothers who have children under age 6, as these mothers either could be treated or would be able to receive treatment in the future. In general, however, the differences between NYC comparison and eligible mothers are minor.

4.5 Population Stability Over Time

The validity of the DID design rests importantly on the assumption that the populations of the eligible and comparison groups do not change over time. I first examine this requirement in my setting by inspecting the change in demographic covariate means from the pre to post period. Table 4 indicates that the underlying demographic composition of the eligible mothers is little changed over the entire time frame. The only variable that changes noticeably in

[44] I construct this binary variable from the ACS variable that reports the amount of pre-tax income (if any) the respondent received during the previous year from various public assistance programs. These include federal/state Supplemental Security Income (SSI) payments to elderly (age 65+), blind, or disabled persons with low incomes; Aid to Families with Dependent Children (AFDC) (now "Temporary Assistance for Needy Families"/TANF); and "General Assistance" (GA). This variable excludes income from charity and separate payments for hospital or other medical care.

Table 2: Outcome Means by Sample and Period

Outcome	Participation Rate		Usual Hours Worked/Week	
	Pre Policy	Post Policy	Pre Policy	Post Policy
Sample	Mean/(SD)/N	Mean/(SD)/N	Mean/(SD)/N	Mean/(SD)/N
Eligible Mothers				
Mothers of NYC 4 Yr Olds	0.62	0.64	21.42	22.80
	(0.49)	(0.48)	(19.36)	(19.37)
	485,659	332,077	485,736	332,077
DID Comparison Groups				
Mothers of MSA 4 Yr Olds	0.67	0.67	23.25	24.27
	(0.47)	(0.47)	(19.47)	(19.30)
	467,657	319,031	467,657	319,031
Mothers of NYC 7 Yr Olds	0.72	0.72	24.87	24.89
	(0.45)	(0.45)	(19.00)	(18.58)
	185,332	142,347	185,332	142,347
DDD Comparison Group				
Mothers of NYC 7 and MSA 4 and 7 Yr Olds	0.69	0.69	23.95	24.84
	(0.46)	(0.46)	(19.26)	(19.04)
	827,887	599,617	827,887	599,617

Notes: Table 2 presents the pre and post period labor force participation rate and average hours worked per week for eligible mothers, MSA comparison mothers, NYC comparison mothers, and DDD comparison mothers using American Community Survey (ACS) data. The pre period covers 2010-2013, while the post period covers 2014-2016. Observations are unweighted.

Table 3: Pre Period Covariate Means for Eligible Mothers and Comparison Mothers

Variable	(1) Eligible Mean/SE	(2) DID 1 Mean/SE	(3) DID 2 Mean/SE	T-test Difference (1)-(2)	T-test Difference (1)-(3)
Age (Mother)	33.71	35.16	38.03	-1.45***	-4.31***
	(0.01)	(0.01)	(0.02)		
# Children	2.38	2.33	2.10	0.05***	0.28***
	(0.00)	(0.00)	(0.00)		
Single	0.34	0.21	0.36	0.13***	-0.02***
	(0.00)	(0.00)	(0.00)		
White	0.40	0.64	0.33	-0.24***	0.07***
	(0.00)	(0.00)	(0.00)		
Hispanic	0.35	0.26	0.35	0.09***	-0.00**
	(0.00)	(0.00)	(0.00)		
College Deg. or Higher	0.31	0.47	0.29	-0.15***	0.03***
	(0.00)	(0.00)	(0.00)		
Diploma/GED Only	0.24	0.21	0.24	0.03***	0.01***
	(0.00)	(0.00)	(0.00)		
In School	0.10	0.06	0.08	0.03***	0.02***
	(0.00)	(0.00)	(0.00)		
Public Assistance Recipient	0.06	0.04	0.07	0.02***	-0.01***
	(0.00)	(0.00)	(0.00)		
N	491545	474762	187210		

Notes: The table above contains pre period (2010 - 2013) covariate means for different samples of mothers. Column 1 includes the sample of eligible mothers (NYC mothers of 4 year olds); Column 2 includes the first DID comparison sample, MSA mothers of 4 year olds; and Column 3 contains the second DID comparison sample, NYC mothers of 7 year olds. Observations are weighted using ACS person weights.

Table 4: Pre and Post Period Covariate Means, Eligible Mothers

	Pre Policy	Post Policy	Difference	p-value
NYC Mothers of 4 Yr Olds				
Age (Mother)	34.11	34.75	-0.64***	
	(6.64)	(6.46)	(0.16)	0.000
# Children	2.43	2.43	0.00	
	(1.37)	(1.38)	(0.03)	0.941
Single	0.30	0.28	0.02*	
	(0.46)	(0.45)	(0.01)	0.038
White	0.43	0.45	-0.02	
	(0.50)	(0.50)	(0.01)	0.072
Hispanic	0.31	0.29	0.02	
	(0.46)	(0.45)	(0.01)	0.078
College Deg. or Higher	0.34	0.41	-0.08***	
	(0.47)	(0.49)	(0.01)	0.000
HS Diploma/GED Only	0.23	0.22	0.01	
	(0.42)	(0.41)	(0.01)	0.231
In School	0.10	0.08	0.02**	
	(0.30)	(0.27)	(0.01)	0.005
Public Assistance Recipient	0.06	0.05	0.01	
	(0.24)	(0.22)	(0.01)	0.167
Observations	3,898	2,709		

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Notes: Table 4 displays covariate means for eligible mothers (NYC Mothers of 4 year olds) in the pre period (2010-2013) and the post period (2014-2016). The table also includes the difference in means between the pre and post period and the p-value from a t-test of the difference. Observations are unweighted.

magnitude is the percent of the population with a college degree or higher, increasing from 33 to 39 percent over the two periods. The eligible group also experiences a 2 percentage point decline in the school attendance rate and share Hispanic, while the share white increases by 2 percentage points (this also occurs for the NYC comparison sample). These differences are all statistically significant.

Tables 13 and 14 in Appendix B.1 compare the covariate means over time for the two DID comparison groups. As with the sample of eligible mothers, both comparison groups experience a 3-4 percentage point increase in the share of the sample with a college degree or higher. There is a small decline over time (2 percentage points) in the share of the MSA sample that is white. Neither of these trends is cause for alarm, as the U.S. population is becoming more diverse and educated over time. Overall, these demographic indicators provide encouraging evidence that the underlying populations for both the treatment and comparison groups are fairly stable over time, at least based on observables.[45]

To confirm that the covariate means are not affected by the treatment offer, I estimate covariate balance regressions. These regressions use the same specification as the DID and DDD models except that the outcome variable is replaced by each covariate (which also is removed from the right hand side). Under the null hypothesis of no changes in the underlying population, the "treatment" parameter α should equal zero. Table 5 presents the results of these regressions, with each column corresponding to a different covariate as the dependent variable. Almost all of the parameter estimates are statistically insignificant and most are close to zero in magnitude.[46] Overall, the results on covariate stability provide reassurance that the UPK treatment does not alter observable covariates in the eligible sample relative to the comparison mothers.

4.6 Evidence of Bias Stability/Parallel Trends

Having considered the demographic composition of the subsamples over time, I now examine the pre period trends to see if the outcome paths evolve in parallel for eligible and comparison mothers. Figure 1 plots mothers' unconditional labor force participation rates in the pre period, stratified by eligibility for UPK. Each panel represents a different model specification, corresponding to one of the comparison groups. The patterns suggest that mothers from both the residency and age samples may be effective comparison groups.

In the left panel, MSA mothers of 4 year olds follow the same trends as the eligible NYC mothers of 4 year olds, with a temporary narrowing of the gap in 2011. A similar, though slightly less pronounced, narrowing occurs using the NYC age comparison group in the middle panel. In both panels, the differences are fairly stable in the two years preceding UPK. The third panel plots the rates for the DDD sample, with NYC mothers of 7 year olds and MSA mothers of 4 and 7 year olds comprising the comparison group. In this model, the MSA mothers of 7 year olds are used to adjust for changes over time in the participation spread between mothers of 4 and 7 year olds. The levels of participation are quite aligned

[45] One still may be concerned that the composition of the population was altered from changes in migration patterns as a result of the policy. I explore this further in Section 6.2.

[46] One exception is the coefficient for disability in the MSA regression, which increases a statistically significant 1 percentage point in the post period.

Table 5: DID Covariate Balance Regressions, by Comparison Group

Dependent Variable	Education	Age Youngest	# Kids	Single	White	Hispanic	Disability	Child Birth Qtr	Age 20-60
MSA 4s									
Post × Elig	−0.03	−0.05	−0.05	−0.01	0.02	−0.00	0.01*	0.02	0.00
	(0.04)	(0.05)	(0.04)	(0.02)	(0.02)	(0.02)	(0.00)	(0.05)	(0.00)
	$p = 0.465$	$p = 0.306$	$p = 0.263$	$p = 0.563$	$p = 0.288$	$p = 0.971$	$p = 0.036$	$p = 0.749$	$p = 0.849$
Observations	15,819	15,819	15,819	15,819	15,819	15,819	15,819	15,819	15,819
NYC 7s									
Post × Elig	−0.00	−0.01	−0.03	−0.00	−0.02	−0.01	0.01	0.00	−0.00
	(0.06)	(0.04)	(0.04)	(0.03)	(0.02)	(0.03)	(0.00)	(0.05)	(0.00)
	$p = 0.976$	$p = 0.710$	$p = 0.544$	$p = 0.856$	$p = 0.459$	$p = 0.616$	$p = 0.209$	$p = 0.959$	$p = 0.810$
Observations	10,495	10,495	10,495	10,495	10,495	10,495	10,495	10,495	10,495
NYC 7s and MSA 4s and 7s									
Post × Elig	−0.01	−0.03	−0.03	−0.00	−0.02	−0.01	0.01	0.01	−0.00
	(0.06)	(0.05)	(0.04)	(0.03)	(0.02)	(0.03)	(0.00)	(0.05)	(0.00)
	$p = 0.892$	$p = 0.547$	$p = 0.509$	$p = 0.882$	$p = 0.444$	$p = 0.609$	$p = 0.200$	$p = 0.910$	$p = 0.849$
Observations	21,940	21,940	21,940	21,940	21,940	21,940	21,940	21,940	21,940

Notes: The table displays estimates from covariate balance regressions for the two DID and the DDD comparison group samples using ACS data from 2010-2016. The regressions also include the mother's age and age^2 as covariates. Standard errors clustered by the Public Use Micro Area (PUMA) are in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

for the mothers of 7 year olds in NYC and the MSA, and they remain fairly flat over time, although they are less parallel.

One difference between the graphs is that the 2010 participation rate gap between eligible and ineligible mothers is much wider for the NYC comparison group than for the MSA comparison group, stemming from a larger participation rate of mothers of 7 year olds versus mothers of 4 year olds (which also seems to be at play in the right panel, though less so). This difference could arise from different effects of the Great Recession on mothers of younger versus school-aged children. [47]

Overall, the comparison mothers' participation rate appears relatively parallel to that of the eligible mothers over the entire pre period, with the exception of 2011. It is likely that this divergence is tied to the NYC preschool market and the availability of public preschool seats, as 2011 was a particularly difficult year to secure a seat in public preschool.[48]

5 Results

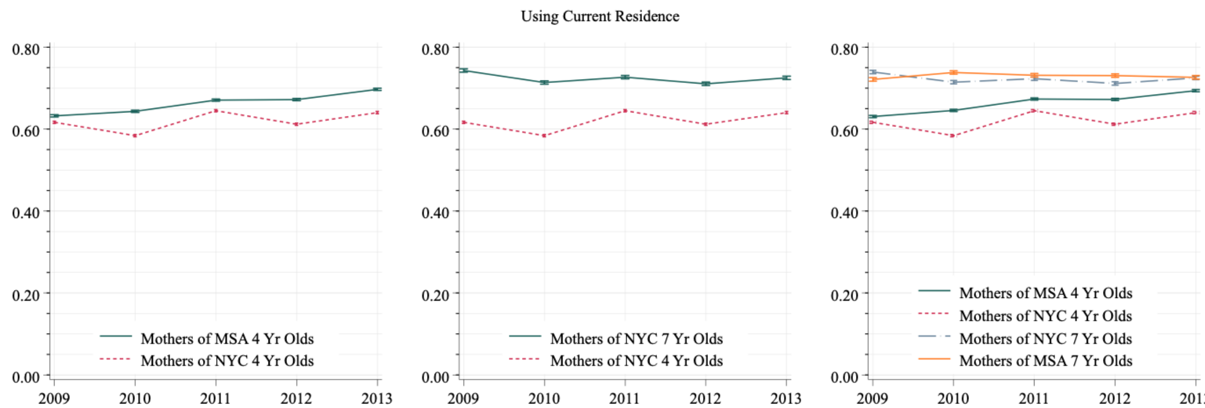
5.1 Difference-in-Differences Model

In this section I present results for the effect of Pre-K For All on mothers' labor supply for the two comparison samples: MSA (non-NYC resident) mothers of 4 year olds and NYC mothers of 7 year olds. The effects on mothers' labor force participation are shown in Tables 6 and 7. The first two columns contain results using data for 2010 through the first year of the policy (2014), while the results in columns (3) and (4) use data for all years (2010-2016). The policy effect is given by the parameter on the interaction of "post" and "eligible" (that is, T_i from Equations 1 - 3), shown in the first row of each table. The columns of interest are (2) and (4), which correspond to the models including covariates. For simplicity, I omit

[47]A potential story could be that the Great Recession created a temporary disjuncture across cohorts in the labor market behavior of mothers of 4 year olds. Suppose, for argument's sake, that during normal economic times, childbirth induces a woman to exit the labor market temporarily and return when her child reaches preschool or kindergarten age, regardless of the birth cohort of her child. Mothers whose time to return to the labor market coincides with a recession may be discouraged and wait for the economy to improve before resuming labor market participation. Yet the mothers of slightly older children would not have faced such discouragement the year prior and so would have reentered the labor market. This differential timing of labor market reentry would create a fissure across cohorts in the participatory behavior of mothers with 4 year olds. To make matters worse, extensions of unemployment insurance programs during the recession would further encourage labor force participation of the already more-attached mothers of slightly older children, as their prior year earnings would qualify them to receive unemployment benefits, while mothers who had been out of the labor force (perhaps due to childbearing) would not have been eligible for such benefits. Moreover, the criteria and timing of these benefit extensions differed across states and so may have affected NY and NJ mothers differently.

[48]In fact, it is likely that this excess demand prompted NYC to conduct a study of 4 year old pre-K, the results of which shed light on seat deficiencies and ultimately led to the universal policy. However, I have not been able to identify any particular source of the relative shortage of preschool seats.

Figure 1: Assessing the Bias Stability Assumption: Mothers' Unconditional Labor Force Participation Rate by UPK Eligibility



Notes: Figure 1 plots mothers' labor force participation rates by UPK eligibility for the years leading up to the policy (2009-2013). Each panel corresponds to a different comparison group sample. The left and center panel use the DID samples, while the right panel uses the DDD sample of mothers. NYC mothers in comparison groups are excluded if they have a child under age 6. Observations are weighted using ACS person weights.

the coefficients on the covariates from these tables, but they are included in Table 15 in Appendix C, which shows the effects for the different comparison groups using all years and covariates.[49]

I find that UPK increases mothers' labor force participation for both comparison groups, with treatment effect estimates ranging from 3.8 to 8.4 percentage points, depending on the time frame considered. The effects are largest in the first year of the policy, with an estimated 6.8 percentage point increase using the MSA comparison and an 8.4 percentage point increase using the NYC age comparison; both estimates are statistically significant at the 1 percent level. When all post-policy years are included, the treatment effect using the MSA and NYC models are approximately equal, estimated as a 3.9 (MSA) and 3.8 (NYC) percentage point increase in labor force participation. The effects are significant at the 1 percent level for the MSA model and the 5 percent level for the NYC model.

I hypothesized (see Section 3.2) that the parallel trends assumption holds conditional on demographic covariates, but perhaps not unconditionally, given the demographic differences in the eligible and ineligible populations (see Table 3). It is reassuring to see from Tables 6 and 7 that the policy effects are quite similar in magnitude whether or not covariates are included, with a maximum effect size spread of 0.4 percentage points. This consistency in estimated magnitudes suggests that the parallel trends assumption may be reasonable even unconditionally. Even so, the inclusion of covariates does appear to remove variation in

[49]The "No Cov." columns correspond to the model estimated without demographic covariates (but including year dummies).

outcomes stemming from underlying population differences between the eligible and comparison groups, given that it increases the estimated effect for the MSA comparison model while dampening the policy effect for the NYC age comparison model.

Tables 16 and 17 in Appendix C show the same regressions for the outcome of usual hours worked per week. The estimates are similar for the first year of the policy (Column 2) regardless of which comparison group is used. For the MSA comparison groups I find that the UPK policy increases usual hours worked per week by 2.4 hours, while the corresponding effect for the NYC comparison group is 2.47 hours. These effects are statistically significant at the 5 percent level. Interestingly, the effects are much larger when data from only the first year of the policy are used. The reason for this larger effect on hours worked may stem from the fact that in the first year of Pre-K For All's implementation, much of the increase in the number of "full-day" pre-K seats came from converting pre-K seats from a half-day (e.g. morning-only or afternoon-only) to full-day schedule. Thus, mothers may have been increasing their work on the intensive margin in response to the increased amount of time that children were spending in pre-K.

Somewhat surprisingly, when later post policy years are included in the estimation, the effects for the two comparison groups diverge. I detect a statistically insignificant increase of 0.7 hours for the MSA DID model, while for the NYC model I estimate a 1.8 hour increase in usual hours worked over the entire post period ($p = 0.019$). Although the sign of the policy effect is positive across specifications and comparison groups, the difference in specifications appears to matter for the magnitude of the effect on hours worked for the entire post period.

Overall, each DID model uncovers a similar effect of the policy on mothers' participation and hours worked despite using different ineligible mothers as comparisons. While this outcome is encouraging, it is not sufficient to ensure the validity of my research design. Indeed, comparing the point estimates in columns (2) and (4) across the two DID models (in Tables 6 and 7) reveals that the estimated policy effect in the first year is 1.6 percentage points smaller for the MSA model than for the NYC age model (8.4 versus 6.8 percentage points). Fortunately, this difference disappears when using all post-policy years. The spread between the estimated effects of the two models in the first year could indicate the presence of omitted variable bias due to group-specific unobservables. For instance, the MSA model may fail to account for unobservables at the state level that coincide with the introduction of the UPK policy, while the age model may mistakenly attribute to the policy any shocks arising in the post period that affect the labor supply decisions of women with 4 year olds but not 7 year olds (or vice versa). The triple differences model described in Section 3.4 may improve upon the DID models by removing state-time level variation in the MSA model or age-time level variation in the age model. I discuss these results in the next section.

5.2 Triple Differences Model

I combine the various contrasts described above into a DDD model to account for group-time specific unobservable effects on the outcomes. The new comparison group comprises NYC and MSA mothers of 4 and 7 year olds and NYC mothers of 7 year olds. Adding a contrast between MSA mothers of 4 and 7 year olds removes from the DID 1 (MSA) estimates any participation-relevant changes over time between mothers of 4 and 7 year olds that are common to both locations. Furthermore, the added contrast between NYC and MSA mothers of 7 year olds helps remove from the DID 2 (age) model estimates any

Table 6: Difference-in-Differences Results, Labor Force Participation

Comparison Group:	2010-2014		2010-2016	
	(1) No Cov.	(2) Covariates	(3) No Cov.	(4) Covariates
Mothers of MSA 4 Yr Olds				
Post × *Eligible*	0.056*	0.063*	0.034*	0.039*
	(0.028)	(0.027)	(0.016)	(0.015)
Post Policy	−0.050*	−0.060*	−0.009	−0.014
	(0.024)	(0.023)	(0.020)	(0.019)
Eligible	−0.051**	−0.040*	−0.051**	−0.037*
	(0.017)	(0.018)	(0.017)	(0.018)
Constant	0.694***	0.155	0.694***	0.116
	(0.015)	(0.127)	(0.015)	(0.101)
R^2	0.004	0.076	0.003	0.076
Adjusted R^2	0.003	0.075	0.003	0.075
Observations	10,184	10,184	15,817	15,817

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Notes: Tables 6 (above) and 7 (below) display results from OLS estimation using ACS data for 2010-2016. Columns correspond to different model specifications. Columns (1)-(2) are for years 2010-2014, while columns (3)-(4) use data for all years. The outcome variable is a binary indicator of a mother's labor force participation. Mothers in the "Eligible" group are current NYC residents with a 4 year old child. "Post" is a binary indicator of the policy period (2014-2016). All specifications contain year dummies and are weighted with ACS person weights. NYC comparison group mothers with children younger than age 6 are excluded from all samples. Covariates include mother's educational attainment (no degree, HS Diploma, some college, bachelor's or higher), age of youngest own child in the household, number of own children in the household, dummy indicators for single, white, Hispanic, independent living difficulty, and child's birth quarter, and mother's age and age^2. Standard errors are clustered by Public Use Micro Area (PUMA).

Table 7: Difference-in-Differences Results, Labor Force Participation

Comparison Group:	2010-2014		2010-2016	
	(1) No Cov.	(2) Covariates	(3) No Cov.	(4) Covariates
Mothers of NYC 7 Yr Olds				
Post × Eligible	0.076**	0.073**	0.036*	0.035*
	(0.025)	(0.022)	(0.017)	(0.017)
Post Policy	−0.063*	−0.063**	0.001	−0.001
	(0.026)	(0.024)	(0.027)	(0.027)
Eligible	−0.099***	0.023	−0.099***	0.016
	(0.014)	(0.024)	(0.014)	(0.021)
Constant	0.735***	0.070	0.735***	0.071
	(0.020)	(0.158)	(0.020)	(0.110)
R^2	0.008	0.103	0.009	0.102
Adjusted R^2	0.008	0.100	0.008	0.100
Observations	6,666	6,666	10,494	10,494

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Notes: See note for Table 6 above.

changes from the pre to post periods that were idiosyncratic to one of the age groups but experienced in both locations. The identifying assumption for the DDD model is that there were no shocks coinciding with the UPK policy that are location and age specific.

Table 8 shows DDD results on maternal labor force participation. As with the DID models, I separately estimate the effects using the first year post policy only (2014) and then using all years post policy (2014-2016). The estimated effect on mothers' participation in the first year of the policy is 12.6 percentage points, which is 5.8 and 4.2 percentage points larger than the estimates using DID models 1 and 2, respectively. This large difference between the DID and DDD results may indicate that the DID estimates suffer from omitted variable bias arising from unobserved time-location and time-age effects. However, these differences also could arise if the DDD model is invalid.

On the other hand, when all years in the post period are used, I estimate that UPK increases maternal labor force participation of eligible NYC mothers by 5 percentage points (although this result is marginally insignificant, with $p = 0.055$). Unlike the estimates for the first year of the policy, the DDD estimate for all post period years is much closer in magnitude to the corresponding DID coefficients (0.039 and 0.038 in the MSA and NYC models, respectively).

Moving on to the effect on hours worked, I find that the DDD design detects a larger effect for both time periods. Table 18 in Appendix C shows an estimated increase in hours worked by 6.2 hours per week for the first year of the policy, with an average increase of 2.9 hours per week over the entire post-policy period. Both of these results are statistically significant.

For both outcomes, the large difference in results between the DID and DDD speci-

Table 8: Triple Differences Results, Labor Force Participation

Comparison Group:	2010-2014		2010-2016	
	(1) No Cov.	(2) Covariates	(3) No Cov.	(4) Covariates
Mothers of NYC 7 and MSA 4 and 7 Yr Olds				
Post Policy × 4 Yr Old=1 × NYC Resident	0.115**	0.121**	0.047	0.050
	(0.043)	(0.041)	(0.026)	(0.027)
Post Policy × 4 Yr Old=1	−0.039	−0.047	−0.012	−0.014
	(0.034)	(0.034)	(0.020)	(0.020)
Post Policy × NYC Resident	−0.059	−0.060	−0.013	−0.012
	(0.035)	(0.033)	(0.023)	(0.024)
4 Yr Old=1 × NYC Resident	−0.037	−0.034	−0.037	−0.036
	(0.020)	(0.020)	(0.020)	(0.020)
Post Policy	−0.005	−0.006	0.004	0.000
	(0.026)	(0.026)	(0.019)	(0.020)
4 Yr Old=1	−0.062***	0.050*	−0.062***	0.034
	(0.015)	(0.022)	(0.015)	(0.020)
NYC Resident	−0.013	−0.007	−0.013	−0.003
	(0.022)	(0.021)	(0.022)	(0.021)
Constant	0.750***	0.068	0.750***	0.055
	(0.016)	(0.118)	(0.016)	(0.096)
R^2	0.008	0.076	0.007	0.076
Adjusted R^2	0.008	0.074	0.007	0.075
Observations	14,032	14,032	21,938	21,938

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Notes: The table displays DDD results from OLS estimation using ACS data for 2010-2016. The outcome variable is a binary indicator of a mother's labor force participation. Columns correspond to different model specifications. Columns (1)-(2) are for years 2010-2014, while columns (3)-(4) use data for all years. The post policy period is 2014-2016. All specifications contain year dummies and are weighted with ACS person weights. NYC comparison group mothers with children younger than age 6 are excluded from all samples. Covariates include mother's educational attainment (no degree, HS Diploma, some college, bachelor's or higher), age of youngest own child in the household, number of own children in the household, dummy indicators for single, white, Hispanic, independent living difficulty, and child's birth quarter, and mother's age and age^2. Standard errors are clustered by Public Use Micro Area (PUMA).

fications raises questions about the validity of the research design. In the next section I investigate the robustness of the models.

6 Robustness Checks

In this section, I consider threats to the internal validity of my estimated policy effects and discuss various sensitivity analyses that explore the stability of the populations and the strength of the parallel trends assumption. Three scenarios stand out as potentially jeopardizing the validity of the DID's identifying assumptions in my context: (1) Agents alter their participation behaviors after the announcement of the policy but before it is implemented; (2) Mothers whose children will be age-eligible in a post-period year change their residence to NYC in order to receive the treatment; or (3) NYC mothers of 7 year olds are affected by unobserved policies or shocks that unravel the parallel trends in the post period. I consider each of these in turn.[50]

6.1 Anticipatory Responses in the Pre Period

Since I design the DID analysis to compare specific age groups (mothers of 4 year olds, mothers of 7 year olds) before and after the policy, the participation behaviors of mothers observed in the pre period (whether eligible or comparison) should not be affected by the policy announcement. In other words, the participation behaviors of pre period eligible mothers should be stable and provide a correct baseline level of participation for the eligible mothers in the post period (treated mothers). The reason is that mothers observed in the pre period with a 4+ year old child have no reason to alter their labor supply before UPK is implemented because their child will be too old to enroll in pre-K once it is introduced (hence, they never can benefit from the policy).[51] As such, all pre period observations included in the analysis sample should not respond to the policy.[52]

Another concern could be that treated agents alter their participation behaviors after the announcement of the policy but before they are actually treated. That is, agents who are

[50]I focus on the DID models for participation but also discuss, where relevant, when the findings have implications for the DDD participation model. Some of the robustness analyses for hours worked (e.g., the migration adjustments) are included in Appendix C; the remainder are available by request.

[51]A mother cannot retroactively manipulate her child's birth date in order to reassign him to a younger cohort and be able to take up the free pre-K offer.

[52]I note that if I were to use NYC mothers of 3 year olds (instead of 7) as a comparison group, I would be more concerned about anticipation changing the behaviors of the comparison mothers. In that model, it would be reasonable to expect mothers of 3 year olds to "re-optimize" once the policy is announced, which could involve changing their labor supply choices even before they can take up the policy. As a result, the underlying population of the "comparison" group in the post period would differ from that in the pre period. Nevertheless, even under this specification, I believe it would be unlikely for mothers of the first treated cohort to adjust their labor market behaviors in 2013 in anticipation of the policy, given the uncertainty of elections and the short time-frame over which Pre-K For All was introduced.

eligible in the post period may have changed their behavior after the policy was announced but when their child was younger than the eligible age. If the outcome at interview is the same as the outcome associated with the mother's initial response to the policy, the DID estimates will capture the immediate effect of the policy on the mother's labor supply. Otherwise, my policy effect estimates will not measure mothers' initial reaction to the policy; rather, they will capture the net effect on mothers' participation that can be detected at the time their children become treated. The effect estimates for the first year of the policy are more likely to be robust to this problem since the time from announcement to implementation was fairly brief. For the second and third year of the policy, treated mothers are more likely to have re-optimized their childcare and labor supply decisions in the years leading up to their child's enrollment in pre-K, and the evolution of these dynamic effects cannot be explored with my DID design.

6.2 Migration-Induced Changes in the Populations

Although the previously discussed age-related anticipatory behaviors should not affect my research design and estimates, migration poses more of a concern because the residence component of eligibility is manipulable by mothers. To satisfy the residency component of UPK eligibility, mothers may respond to the policy in two ways. First, MSA resident mothers whose child will turn 4 in the future (and so eventually will satisfy the age-eligibility criteria) may decide to move to NYC to take advantage of free pre-K. Second, NYC resident mothers who had planned to leave the City before their child turns 4 may instead decide to stay in NYC since they can receive free pre-K.

While it would be reasonable to regard any changes in migration as a relevant effect of the policy, changes in migration would complicate the interpretation of the policy effect. In particular, the presence of these migratory changes could undermine the assumption that the underlying populations are stable over time. Empirically, the introduction of the policy in 2014 does not appear to induce migratory flows of 4 year olds into NYC from the rest of the MSA. However, I find evidence that the policy reduces the outflow of mothers of 3 and 4 year olds from NYC.

To account for these migratory changes, I conduct robustness analyses that estimate the models using a sample selected based on residency one year prior, rather than based on current residence. The value of this approach is that mothers decide their prior year residency before the policy is introduced (at least for the first year of the policy), so the composition of the sample should not be affected by UPK. For the MSA comparison, all mothers of 4 year olds who lived in the MSA (NYC inclusive) one year before are included in the sample; those who resided in NYC proper last year are considered "eligible" for the policy. For the NYC comparison, those who lived in NYC proper last year are selected for the sample, and the age criteria is unchanged. With this approach, a UPK-induced change in out-migration can be ignored because the overall participation rate of eligible mothers is unaffected by the breakdown of NYC stayers and leavers.

Figure 2 in Appendix C.1 examines the trends in the pre period for the sample selected based on the mother's residence one year prior. The trends are not noticeably different than when current year residence is used to select the sample. However, regression estimates presented in Table 9 suggest that using prior year residence may sidestep migration-induced changes to the underlying populations. For the MSA DID specification, the policy effect estimates for the entire post-policy period are a statistically significant 3.3 percentage points using prior year residence ($p = 0.019$), compared to 3.9 percentage points using current residence. Similarly, the estimated effect for the NYC comparison DID is 3.4 percentage points ($p = 0.045$) using prior year residence, marginally less than the 3.8 percentage point increase in participation estimated using current residence.

6.3 Placebo Policies

To investigate further the robustness of the model assumptions, I examine whether effects can be detected by placebo policies. These exercises can help detect if there are other unobserved policies or shocks that could confound the estimated effects. I take two approaches. First, I vary the introduction of the policy to earlier years (2010-2013) and re-estimate the models using data through 2013 with this placebo "post policy" year. Second, I vary the construction of the "eligibility" variable by estimating the model with 6-9 year olds as the "treated" age groups (using all years, 2010-2016). The logic of these tests is that the true effect should be zero: The policy should not have an effect before it was introduced, and the policy should not affect age groups that are beyond the preschool age. Any detection of nonzero effects in the former case would call into question whether the ineligible mothers provide an appropriate comparison for eligible mothers in the non-treated state of the world. Estimates of nonzero effects in the latter case would suggest that the comparison groups are affected by the UPK policy, which could indicate the presence of other policies or shocks that occur simultaneously with UPK. Evidence of such unobserved confounding factors would suggest that the estimates be interpreted with caution.

Tables 11 and 12 present the results from these placebo tests. The first panel of Table 11 shows the coefficient on the interaction for each of the placebo years. This regression is equivalent to the specification for DID 1 (Equation 2), except that instead of defining the post policy period to begin in 2014, I define it to begin with the year indicated by the column. The coefficient represents the differential effect of the placebo policy on labor force participation for NYC mothers of 4 year olds relative to their MSA counterparts. For each year, the coefficients are small in magnitude, ranging from -0.01 to -0.02, and are statistically indistinguishable from zero.

The second panel in Table 11 presents the policy effect coefficient using the same placebo definition of the post-policy period as panel 1, but for the DID 2 specification (Equation 3) and sample. The results are similar for this comparison group: None of the estimates is statistically distinguishable from zero. Furthermore, the coefficients are small in magnitude, estimated to be -0.01 in 2010, 0.01 in 2011 and 2012, and 0.02 in 2013. In summary, when using the placebo policy years I do not detect any effects that are statistically different from zero, suggesting that, at least in the pre-policy years, the participation of comparison mothers did not trend differently from that of the eligible mothers.

Table 12 presents the findings from a similar placebo experiment, but instead of changing the policy year, I alter the age group receiving the "treatment." Each column corresponds to a modified Equation 2 (MSA DID) in which the eligibility variable is defined using the age given in the column title. For all of these regressions, I only use mothers whose youngest child is the placebo age; since NYC mothers with older children also may have a 4 year old that was treated, they need to be removed from the sample when testing the effect of the placebo age eligibility. For comparison, the leftmost column shows the treatment effect using

Table 9: DID Results, Labor Force Participation (with Migration Comparison)

Comparison Group	MSA 4 Yr Olds		NYC 7 Yr Olds	
	Current Residence	Residence 1 Yr Prior	Current Residence	Residence 1 Yr Prior
Main Effects:				
Post Policy × Eligible	0.039*	0.035*	0.035*	0.034*
	(0.015)	(0.015)	(0.017)	(0.017)
Post Policy	−0.014	−0.010	−0.001	−0.003
	(0.019)	(0.019)	(0.027)	(0.026)
Eligible	−0.037*	−0.038*	0.016	0.019
	(0.018)	(0.017)	(0.021)	(0.021)
Constant	0.116	0.081	0.071	0.033
	(0.101)	(0.099)	(0.110)	(0.106)
R^2	0.076	0.077	0.102	0.101
Adjusted R^2	0.075	0.076	0.100	0.098
Observations	15,817	15,747	10,494	10,675

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Notes: Table 9 displays results from OLS estimation using ACS data for 2010-2016. columns correspond to the two DID comparison groups of mothers and whether the sample was selected using the mother's current residence or residence one year prior. The outcome variable is a binary indicator of a mother's labor force participation. Mothers in the "Eligible" group are NYC residents with a 4 year old child. "Post" is a binary indicator of the policy period (2014-2016). All specifications contain year dummies and are weighted with ACS person weights. NYC comparison group mothers with children younger than age 6 are excluded from all samples. Covariates include mother's educational attainment (no degree, HS Diploma, some college, bachelor's or higher), age of youngest own child in the household, number of own children in the household, dummy indicators for single, white, Hispanic, independent living difficulty, and child's birth quarter, and mother's age and age^2. Standard errors are clustered by Public Use Micro Area (PUMA).

Table 10: Triple Differences Results, Labor Force Participation (Migration Adjusted)

Comparison Group:	2010-2014	2010-2016
Post Policy × 4 Yr Old=1 × Mom Lived in NYC Last Yr=1	0.124**	0.047
	(0.043)	(0.026)
Post Policy × 4 Yr Old=1	−0.049	−0.011
	(0.035)	(0.019)
Post Policy × Mom Lived in NYC Last Yr=1	−0.061	−0.013
	(0.033)	(0.023)
4 Yr Old=1 × Mom Lived in NYC Last Yr=1	−0.034	−0.036
	(0.020)	(0.020)
Post Policy	−0.005	0.002
	(0.027)	(0.020)
4 Yr Old=1	0.053*	0.038
	(0.022)	(0.021)
Mom Lived in NYC Last Yr=1	−0.008	−0.004
	(0.021)	(0.021)
Constant	0.043	0.034
	(0.115)	(0.094)
R^2	0.076	0.077
Adjusted R^2	0.074	0.076
Observations	13,950	21,827

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Notes: The table displays results from OLS estimation using ACS data for 2010-2016. columns correspond to the two comparison groups of mothers. The outcome variable is a binary indicator of a mother's labor force participation. Mothers in the "Eligible" group are current NYC residents with a 4 year old child. "Post" is a binary indicator of the policy period (2014-2016). All specifications contain year dummies and are weighted with ACS person weights. NYC comparison group mothers with children younger than age 6 are excluded from all samples. Covariates include mother's educational attainment (no degree, HS Diploma, some college, bachelor's or higher), age of youngest own child in the household, number of own children in the household, dummy indicators for single, white, Hispanic, independent living difficulty, and child's birth quarter, and mother's age and age^2. Standard errors are clustered by Public Use Micro Area (PUMA).

Table 11: DID Using Placebo Policy Years, Labor Force Participation

Panel 1

Comparison Group:	Placebo Policy Year			
Mothers of MSA 4 Yr Olds	2010	2011	2012	2013
(Placebo) Post × Elig	−0.03	−0.02	−0.03	−0.02
	(0.02)	(0.02)	(0.02)	(0.02)
	$p = 0.150$	$p = 0.397$	$p = 0.160$	$p = 0.344$
Observations	12,391	12,391	12,391	12,391

Panel 2

Comparison Group:	Placebo Policy Year			
Mothers of NYC 7 Yr Olds	2010	2011	2012	2013
(Placebo) Post × Elig	−0.01	0.02	0.01	0.01
	(0.02)	(0.03)	(0.03)	(0.03)
	$p = 0.534$	$p = 0.523$	$p = 0.687$	$p = 0.735$
Observations	7,879	7,879	7,879	7,879

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Notes: Table 11 displays "placebo regression" DID results from OLS estimation using ACS data for the pre-policy period (2008-2013). Each panel corresponds to a comparison group (MSA mothers of 4 year olds and NYC mothers of 7 year olds, respectively). This baseline sample includes only mothers whose 4 or 7 year old is their youngest child. Columns correspond to the placebo year of the policy introduction. Although the regressions are estimated with the full set of covariates, for simplicity, the table includes only the policy effect parameter and its standard error and p-value. The outcome variable is a binary indicator of a mother's labor force participation. Mothers in the "Eligible" group are current NYC residents with a 4 year old child. "Post" is a binary indicator of the placebo policy period. All specifications contain year dummies and are weighted with ACS person weights. Standard errors are clustered by Public Use Micro Area (PUMA).

the actual treatment eligibility age of 4 years old. Panel 1 presents results for the first year of the UPK policy (2010-2014), while Panel 2 presents results using 2010-2016.

These coefficients are interpreted as the effect of the placebo age treatment on NYC mothers relative to their MSA counterparts. All of the placebo policy coefficients in the first panel of Table 12 are statistically insignificant, although that could be influenced by the smaller sample size induced by removal of mothers with younger children. The magnitude of the coefficients suggests that mothers of 7-9 year olds in NYC experienced different trends in labor force participation relative to their MSA counterparts in the first year of the policy. In particular, the participation rate of NYC mothers of 7 year olds was 6 percentage points ($p = 0.09$) lower than their MSA counterparts after UPK was introduced, even though neither group of mothers should have been affected by UPK. These findings may indicate the presence of unobserved contemporaneous shocks that affect mothers of 7 year olds differently depending on their residence, or that the participation behaviors of these mothers simply trend differently from one another. Either of these reasons would be problematic for the DDD, because it uses MSA mothers of 7 year olds to remove variation over time in the participation gap between mothers of 4 and 7 year olds in NYC. If mothers of 7 year olds in NYC and the MSA respond differently to unobserved shocks or do not move in parallel, then the added contrast in the DDD actually may yield biased results. Indeed, the effect of -6 percentage points detected for the 7 year old placebo age policy may explain the divergence between the DDD and DID results for the first year of the policy. Given these findings, it may be reasonable to conclude that the DDD results overestimate the policy effect in the first year.

Panel 2 of Table 12 estimates the regression with all post policy years (2010-2016). The coefficient magnitude for mothers of 7 year olds is much smaller (-0.01) and the corresponding p-value is 0.73. The results are also small in magnitude for mothers of 6 year olds, but for mothers of 8 year olds, the coefficient is -0.08 and statistically significant. The finding of large and statistically significant effects for mothers of 8 year olds may indicate that these mothers are less comparable across the two geographic areas, perhaps driven by New York and New Jersey's different unemployment insurance or other policy responses during the Great Recession (see the discussion in Section 4.6). However, given that there do not appear to be differential effects by residency for mothers of 7 year olds when all post-policy years are used, the corresponding DDD results may be more reliable than those for the first year of the policy.

6.4 No Cohort Effects

A DID analysis using a fixed age group (e.g. 4 year olds) over multiple years requires the assumption that expected outcomes are independent of cohort membership. For cohorts of 4 year olds in the years preceding UPK and the first few years after its policy introduction, it is clear that a mother would have been unable to time the birth of her child to ensure that he would turn 4 in the post-policy period. Nevertheless, one might be concerned that mothers who chose to have children during the Great Recession have fundamentally different preferences—in ways that matter for labor supply behaviors—than their counterparts who chose to have children in the expansion period before the recession (pre-2008) or in the post-recession expansion period (e.g. after 2011).

While this concern is reasonable, I believe there exists a sub-sample of cohorts that

Table 12: DID Using Placebo Age Eligibility, Labor Force Participation

Panel 1: First Year of Policy

Comparison Group:	True Age Eligibility	Placebo Age Eligibility			
MSA Mothers	4	6	7	8	9
Post × (Placebo) Elig	0.07	−0.01	−0.06	−0.04	−0.04
	(0.04)	(0.04)	(0.03)	(0.04)	(0.05)
	$p = 0.070$	$p = 0.855$	$p = 0.089$	$p = 0.382$	$p = 0.337$
Observations	4,133	3,776	3,635	3,533	3,367

Panel 2: All Years Post Policy

Comparison Group:	True Age Eligibility	Placebo Age Eligibility			
MSA Mothers	4	6	7	8	9
Post × (Placebo) Elig	0.06*	0.01	−0.01	−0.08**	−0.04
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
	$p = 0.012$	$p = 0.657$	$p = 0.733$	$p = 0.004$	$p = 0.235$
Observations	6,345	6,013	5,786	5,704	5,522

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Notes: Table 12 displays "placebo regression" DID results from OLS estimation using ACS data for 2010-2016. Panel 1 uses only the first year of the post-policy period (2014) while Panel 2 uses all post-policy years. Columns correspond to different placebo ages for eligibility. Although the regressions are estimated with the full set of covariates, for simplicity, the table includes only the policy effect parameter and its standard error and p-value. The outcome variable is a binary indicator of a mother's labor force participation. Mothers in the "Eligible" group are current NYC residents with a child whose age corresponds to the age in the column. "Post" is a binary indicator of the policy period. All specifications contain year dummies and are weighted with ACS person weights. Standard errors are clustered by Public Use Micro Area (PUMA).

would not be subject to these concerns. Specifically, at the time of making their fertility decisions, mothers of the cohort born in 2010 (the first cohort of UPK-treated 4 year olds) and 2011 would have experienced similar economic circumstances as most mothers of the 2008 and 2009 cohorts, since the Great Recession and its effects spanned that period. It seems reasonable to conclude that cohort effects should not be problematic for analyses that use data from two years before and two years after the policy.[53]

To test this hypothesis, I re-estimate the MSA regressions restricting the analysis to these years. The estimated effects of UPK on mothers' participation increase noticeably in magnitude, with a coefficient of 0.052 (versus 0.039) using current residence and a coefficient of 0.045 using prior year residence (versus 0.033). These effects are also statistically significant at the 1 percent level.

6.5 Summary

The results from the preceding robustness analyses provide further evidence that UPK causes women's labor supply to increase. Although adjustments for migration cause a reduction in the magnitude of the participation effect, when all post-policy years are used the DID coefficients still detect an economically significant increase in participation of 3.3 to 3.5 percentage points. The placebo tests further support these findings. The corresponding results from the DDD analyses are slightly larger in magnitude at 4.5 percentage points (although they are not statistically significant).

On the other hand, results using the first year of the policy may be less reliable. Analyses using placebo age policies indicate that estimates using an NYC-MSA comparison of mothers of 7 year olds may suffer from bias due to diverging participation paths in the first year of the policy. This finding, along with the sizable gap in magnitudes between the DDD and DID results, together suggest that the DDD results for the first year of the policy should be interpreted with caution. Nevertheless, the sign and magnitude of all results are consistently positive and indicate that UPK may enable NYC mothers to overcome barriers to their labor force participation.

7 Concluding Remarks

The increase in women's labor force participation was a large driver of U.S. economic growth in the second half of the 20th century. In the last two decades, however, this participation has fallen and U.S. women are choosing to have fewer children. As policymakers and researchers try to explain these concurrent phenomena, the cost of having children (including the direct costs of childcare and the indirect costs to women's professional advancement) has been

[53] That is, 2012-2015 ACS data for mothers of 4 year olds/children born in 2008-2011. Note that I do not believe that this logic would hold for the NYC comparison mothers. The reason is that restricting the analysis sample to ACS years 2012-2015 would reduce the NYC comparison sample to mothers whose children were born from 2005 to 2008. Given that these mothers would have made their fertility decisions during the economic boom that was occurring in 2004-2007, they likely are very different than the mothers of 4 year olds in the same year, who made their fertility decisions during poor economic times.

suggested as a possible contributing mechanism. Indeed, as culture and societal expectations change over time, it seems reasonable that policies affecting how women allocate their time and monetary resources between work and family could also induce generational shifts in labor and fertility behaviors.

In this paper, I offer new evidence relating labor supply decisions to childcare costs by investigating how the introduction of universal pre-K affects mothers' labor force participation on the extensive and intensive margins. Even the most conservative estimates detect a 3.3 percentage point increase in mothers' labor force participation in response to the policy. In the first year of the policy, the DID models estimate an effect ranging from 6.8 to 8.4 percentage points. These results are statistically significant.

It is important to recognize the limitations of the conclusions that can be drawn from this study. First, the effects of the policy may be context dependent. New York City residents may have preferences and behaviors that have limited generalizability to other settings, and the design and implementation of a new universal pre-K policy likely would differ if pursued in other cities or at a statewide level, even if it was modeled after Pre-K For All. Furthermore, the baseline levels of preschool access and preschool enrollment may vary across locations and would affect the room for growth in preschool enrollment, the channel through which I expect to see an effect on maternal labor supply; baseline levels of labor force participation may also vary geographically. The implication is that locations with lower (higher) baseline preschool enrollment and/or access and lower (higher) maternal labor participation potentially could realize greater (smaller) effects on the outcomes. Nevertheless, the results from my study offer an important first step in documenting how mothers in the post-2000 era alter their labor supply in response to the offer of free 4 year old preschool.

References

Battistin, E. and Sianesi, B. (2011). Misclassified Treatment Status and Treatment Effects: An Applications to Returns to Education in the United Kingdom. *The Review of Economics and Statistics*, 93(2):495–509.

Cascio, E. U. (2009). Maternal Labor Supply and the Introduction of Kindergartens into American Public Schools. *Journal of Human Resources*, 44(1):140–170.

Cascio, E. U., Haider, S. J., and Nielsen, H. S. (2015). The effectiveness of policies that promote labor force participation of women with children: A collection of national studies. *Labour Economics*, 36:64–71.

Crawford, S. P., Lader, M.-C., and Smith, M. (2015). On the Road to "Pre-K for All": The Launch of UPK in New York City. Technical report, The Berkman Center for Internet & Society Research.

Duncan, G. J., Ludwig, J., and Magnuson, K. A. (2007). Reducing Poverty through Preschool Interventions. *The Future of Children*, 17(2):143–160.

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

Fitzpatrick, M. D. (2010). Preschoolers Enrolled and Mothers at Work? The Effects of Universal Prekindergarten. *Journal of Labor Economics*, 28(1):51–85.

Fitzpatrick, M. D. (2012). Revising Our Thinking About the Relationship Between Maternal Labor Supply and Preschool. *The Journal of Human Resources*, 47(3):583–612.

Gelbach, J. B. (2002). Public Schooling for Young Children and Maternal Labor Supply. *The American Economic Review*, 92(1):307–322.

Heckman, J. J. (2006). Skill Formation and the Economics of Investing in Disadvantaged Children. *Science*, 312(5782):1900–1902.

Herbst, C. M. (2017). Universal Child Care, Maternal Employment, and Children's Long-Run Outcomes: Evidence from the US Lanham Act of 1940. *Journal of Labor Economics*, 35(2):519–564.

Lewbel, A. (2007). Estimation of Average Treatment Effects with Misclassification. *Econometrica*, 75(2):537–551.

Olivetti, C. and Petrongolo, B. (2017). The Economic Consequences of Family Policies: Lessons from a Century of Legislation in High-Income Countries. *Journal of Economic Perspectives*, 31(1):205–230.

Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., and Sobek, M. (2019). *IPUMS USA: Version 9.0 [dataset]*. IPUMS, Minneapolis, MN:.

Voyles, L. and Ruess, M. (2019). Pre-K in American Cities. Technical report, CityHealth and The National Institute for Early Education Research, Washington, D.C.

A Measurement Error

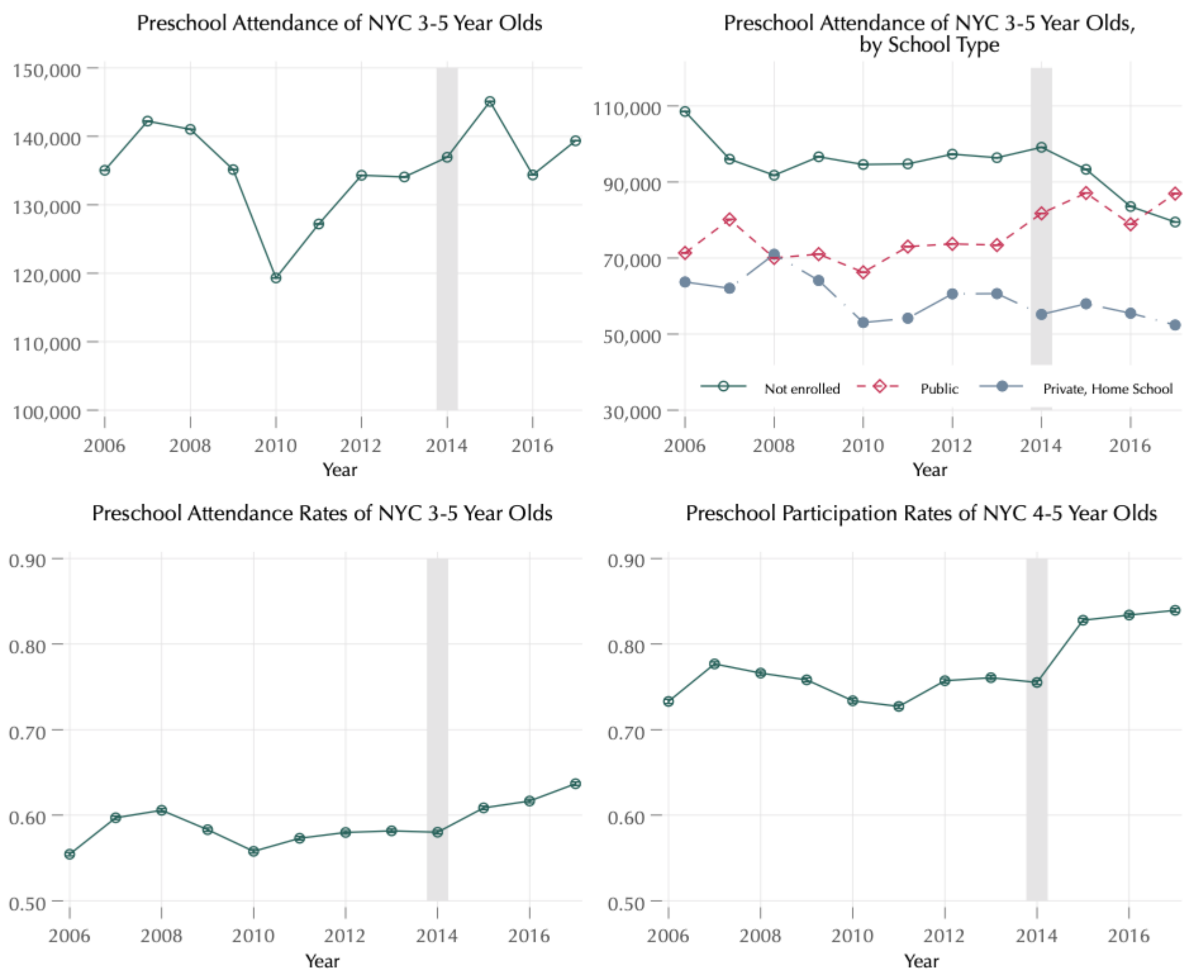
I consider various alternative approaches to understand and correct the measurement error described in Section 4.3. These are works in progress and should be understood to be preliminary and incomplete.

The first exercise is a simple Monte Carlo simulation, wherein I define a data generating process with uniformly distributed birth dates and interview dates, construct the true academic year and age eligibility variables, and then estimate regressions under two sample selection criteria: (1) Construct the sample as if I observed all true data; and (2) Construct the sample using only the proxy variables to determine eligibility criteria (specifically, selecting observed 4 year olds only). In initial simulations, I find the parameter estimates to be attenuated to approximately 2/3 the true parameter value. More work can be done to assess the robustness of these simulations, but they give an initial impression that my measurement error may cause attenuation, despite the nontraditional error in my setting.

The second approach is inspired by multiple imputation. I am still in the early stages of investigation for this exercise, and more work must be done to verify that the intuition is internally valid. Nevertheless, I present the general idea here. First, I categorize children into cells based on 3 pieces of observable information: The calendar year of interview, the birth quarter, and the observed age at interview. By Bayes' Rule, it is possible to compute the probability that an individual in a given cell should be included in the sample for a given academic year. I use simulations to compute these probabilities (as before, creating observations with uniformly distributed birth dates and interview dates) and then assign them to each child's record based on his cell. Then, for each ACS record that could be included in the sample, I use these probabilities to draw either a zero or one from a Bernoulli distribution with the corresponding probability. Observations assigned a one are selected for inclusion in the sample, and then I estimate the DID analysis described in the main text. I repeat this sampling and estimation process many times and compute the average estimate of the policy and its corresponding standard error. Since it is likely fairly common that researchers encounter data limitations for time or eligibility/treatment variables that preclude DID analysis (or are simply ignored), I am considering expanding this exercise to a separate project. The objective would be to develop a methodology for addressing these types of mismeasurement (e.g. time period or eligibility assignment variables) in DID settings.

Finally, I consider a variation on the previous exercise. This approach is also inspired by multiple imputation. Instead of using the aforementioned observables to compute the probabilities that an observation should be included in the sample for each academic year, I would use the observables to compute the probability that a given observation is misclassified in a given case. For example, a case could be that an observation is incorrectly included in the sample when interviewed before receiving the treatment. The cases can be thought of as missing dummy variables that need to be constructed via multiple imputation. The inclusion of these dummy variable for the cases would operate as nonparametric controls for observations that are likely incorrectly included in the sample. The benefit of this approach is that it would allow estimation of the models using all observations, and interpretation would be straightforward because the dummies would control for the "bad" variation contributed by these incorrectly included observation. I would implement the approach as follows: For each record, I would use the probability for each case to draw a zero or one that the case was satisfied. This would create a set of dummy variables for all the cases of incorrect sample inclusion. I then would estimate the regression including these case dummies. Finally, this process would be repeated many times, and the average effect computed.

B Descriptive Evidence



Notes: Sample includes only residents of New York City unless otherwise stated.

Source: Author's calculations using American Community Survey data (U.S. Census Bureau).

B.1 Summary Statistics for Comparison Groups

C Results

Table 13: Pre and Post Period Covariate Means, MSA Mothers of 4 Year Olds

Mothers of MSA 4 Yr Olds	Pre Policy	Post Policy	Difference	p-value
Age (Mother)	35.53	35.49	0.04	
	(5.77)	(5.87)	(0.14)	0.779
# Children	2.33	2.33	0.01	
	(1.12)	(1.14)	(0.03)	0.808
Single	0.18	0.18	−0.00	
	(0.38)	(0.39)	(0.01)	0.841
White	0.67	0.66	0.01	
	(0.47)	(0.47)	(0.01)	0.224
Hispanic	0.21	0.21	0.01	
	(0.41)	(0.40)	(0.01)	0.363
College Deg. or Higher	0.52	0.55	−0.03**	
	(0.50)	(0.50)	(0.01)	0.007
HS Diploma/GED Only	0.18	0.16	0.02*	
	(0.39)	(0.37)	(0.01)	0.021
In School	0.06	0.06	−0.00	
	(0.24)	(0.25)	(0.01)	0.759
Public Assistance Recipient	0.03	0.03	0.01	
	(0.18)	(0.17)	(0.00)	0.188
Observations	4,403	2,999		

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Notes: The table displays covariate means for the MSA mothers of 4 year olds comparison group in the pre period (2010-2013) and the post period (2014-2016). The table also includes the difference in means between the pre and post period and the p-value from a t-test of the difference. Observations are unweighted.

Table 14: Pre and Post Period Covariate Means, NYC Mothers of 7 Year Olds

Mothers of NYC 7 Yr Olds	Pre Policy	Post Policy	Difference	p-value
Age (Mother)	38.66	38.95	-0.29	
	(6.70)	(6.63)	(0.26)	0.274
# Children	2.11	2.15	-0.04	
	(1.13)	(1.06)	(0.04)	0.400
Single	0.31	0.30	0.00	
	(0.46)	(0.46)	(0.02)	0.800
White	0.37	0.39	-0.03	
	(0.48)	(0.49)	(0.02)	0.171
Hispanic	0.32	0.32	-0.00	
	(0.46)	(0.47)	(0.02)	0.984
College Deg. or Higher	0.32	0.36	-0.04*	
	(0.46)	(0.48)	(0.02)	0.023
HS Diploma/GED Only	0.23	0.24	-0.01	
	(0.42)	(0.42)	(0.02)	0.714
In School	0.08	0.06	0.02*	
	(0.28)	(0.24)	(0.01)	0.032
Public Assistance Recipient	0.06	0.04	0.02**	
	(0.24)	(0.19)	(0.01)	0.007
Observations	1,471	1,178		

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Notes: The table displays covariate means for the NYC mothers of 7 year olds comparison group in the pre period (2010-2013) and the post period (2014-2016). The table also includes the difference in means between the pre and post period and the p-value from a t-test of the difference. Observations are unweighted.

Table 15: Difference-in-Differences Results, Labor Force Participation

By Comparison Group	MSA 4 Yr Olds	NYC 7 Yr Olds
Main Effects:		
Post Policy × Eligible	0.039*	0.035*
	(0.015)	(0.017)
Post Policy	−0.014	−0.001
	(0.019)	(0.027)
Eligible	−0.037*	0.016
	(0.018)	(0.021)
Controls:		
HS Diploma/GED	0.107***	0.112***
	(0.015)	(0.015)
Some College	0.206***	0.206***
	(0.018)	(0.017)
Bachelor's or Higher	0.250***	0.291***
	(0.019)	(0.021)
Youngest Child's Age	0.019***	0.023***
	(0.004)	(0.005)
# Own Children	−0.030***	−0.025***
	(0.008)	(0.006)
Single	0.166***	0.151***
	(0.013)	(0.014)
White	−0.025	−0.058***
	(0.014)	(0.013)
Hispanic	0.005	0.010
	(0.015)	(0.016)
Has Independent Living Difficulty	−0.468***	−0.400***
	(0.040)	(0.042)
Age (Mother)	0.014*	0.008
	(0.007)	(0.007)
Age^2	−0.000	−0.000
	(0.000)	(0.000)
Aged 20-60	0.075	0.121
	(0.074)	(0.088)
Constant	0.116	0.071
	(0.101)	(0.110)
R^2	0.076	0.102
Adjusted R^2	0.075	0.100
Observations	15,817	10,494

Notes: The table displays results from OLS estimation using ACS data for 2010-2016. columns correspond to the two DID comparison groups of mothers. The outcome variable is a binary indicator of a mother's labor force participation. Mothers in the "Eligible" group are current NYC residents with a 4 year old child. "Post" is a binary indicator of the policy period (2014-2016). All specifications contain year dummies and are weighted with ACS person weights. NYC comparison group mothers with children younger than age 6 are excluded from all samples. Covariates include mother's educational attainment (no degree, HS Diploma, some college, bachelor's or higher), age of youngest own child in the household, number of own children in the household, dummy indicators for single, white, Hispanic, independent living difficulty, and child's birth quarter, and mother's age and age^2. Standard errors are clustered by Public Use Micro Area (PUMA).

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 16: DID Results, Usual Hours Worked Per Week

Comparison Group:	2010-2014		2010-2016	
	(1) No Cov.	(2) Covariates	(3) No Cov.	(4) Covariates
Mothers of MSA 4 Yr Olds				
Post × Eligible	2.533*	2.570*	0.632	0.623
	(1.141)	(1.087)	(0.700)	(0.683)
Post Policy	−2.051*	−2.356*	−0.000	−0.361
	(0.983)	(0.966)	(0.831)	(0.817)
Eligible	−1.974**	−0.970	−1.974**	−0.774
	(0.727)	(0.637)	(0.727)	(0.631)
Constant	24.950***	2.103	24.950***	−3.012
	(0.648)	(4.684)	(0.648)	(3.601)
R^2	0.005	0.095	0.005	0.095
Adjusted R^2	0.004	0.093	0.005	0.094
Observations	10,185	10,185	15,819	15,819

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 17: DID Results, Usual Hours Worked Per Week

Comparison Group:	2010-2014		2010-2016	
	(1) No Cov.	(2) Covariates	(3) No Cov.	(4) Covariates
Mothers of NYC 7 Yr Olds				
Post × Eligible	2.708*	2.473*	1.941*	1.796*
	(1.264)	(1.156)	(0.827)	(0.783)
Post Policy	−2.039	−2.085	−1.321	−1.598
	(1.120)	(1.047)	(1.184)	(1.154)
Eligible	−3.545***	0.675	−3.545***	0.754
	(0.692)	(1.166)	(0.692)	(0.980)
Constant	26.335***	−0.490	26.335***	−3.839
	(0.847)	(5.683)	(0.847)	(3.874)
R^2	0.009	0.118	0.007	0.120
Adjusted R^2	0.008	0.116	0.007	0.118
Observations	6,667	6,667	10,495	10,495

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 18: DDD Results, Usual Hours Worked Per Week

Comparison Group:	2010-2014		2010-2016	
	(1) No Cov.	(2) Covariates	(3) No Cov.	(4) Covariates
Mothers of NYC 7 and MSA 4 and 7 Yr Olds				
Post Policy × 4 Yr Old=1 × NYC Resident	6.243**	6.342***	2.869*	2.865*
	(1.874)	(1.783)	(1.146)	(1.119)
Post Policy × 4 Yr Old=1	−3.531*	−3.790**	−0.922	−0.976
	(1.390)	(1.339)	(0.799)	(0.798)
Post Policy × NYC Resident	−3.707*	−3.785**	−2.235*	−2.239*
	(1.464)	(1.436)	(1.050)	(1.024)
4 Yr Old=1 × NYC Resident	−2.435*	−2.173*	−2.435*	−2.226*
	(0.964)	(0.876)	(0.964)	(0.871)
Post Policy	1.665	1.627	0.959	0.662
	(1.029)	(1.047)	(0.816)	(0.858)
4 Yr Old=1	−1.114	2.314*	−1.114	2.209*
	(0.674)	(0.921)	(0.674)	(0.856)
NYC Resident	0.458	1.108	0.458	1.345
	(0.921)	(0.899)	(0.921)	(0.888)
Constant	25.880***	−3.535	25.880***	−6.940*
	(0.743)	(4.354)	(0.743)	(3.312)
R^2	0.007	0.089	0.006	0.090
Adjusted R^2	0.007	0.088	0.006	0.089
Observations	14,033	14,033	21,940	21,940

Notes: The table displays results from OLS estimation using ACS data for 2010-2016. The outcome variable is the mother's usual hours worked per week. Columns correspond to different model specifications. Columns (1)-(2) are for years 2010-2014, while columns (3)-(4) use data for all years. Mothers in the "Eligible" group are current NYC residents with a 4 year old child. "Post" is a binary indicator of the policy period (2014-2016). All specifications contain year dummies and are weighted with ACS person weights. Standard errors are clustered by Public Use Micro Area (PUMA).

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

C.1 Robustness: Correcting for Migration

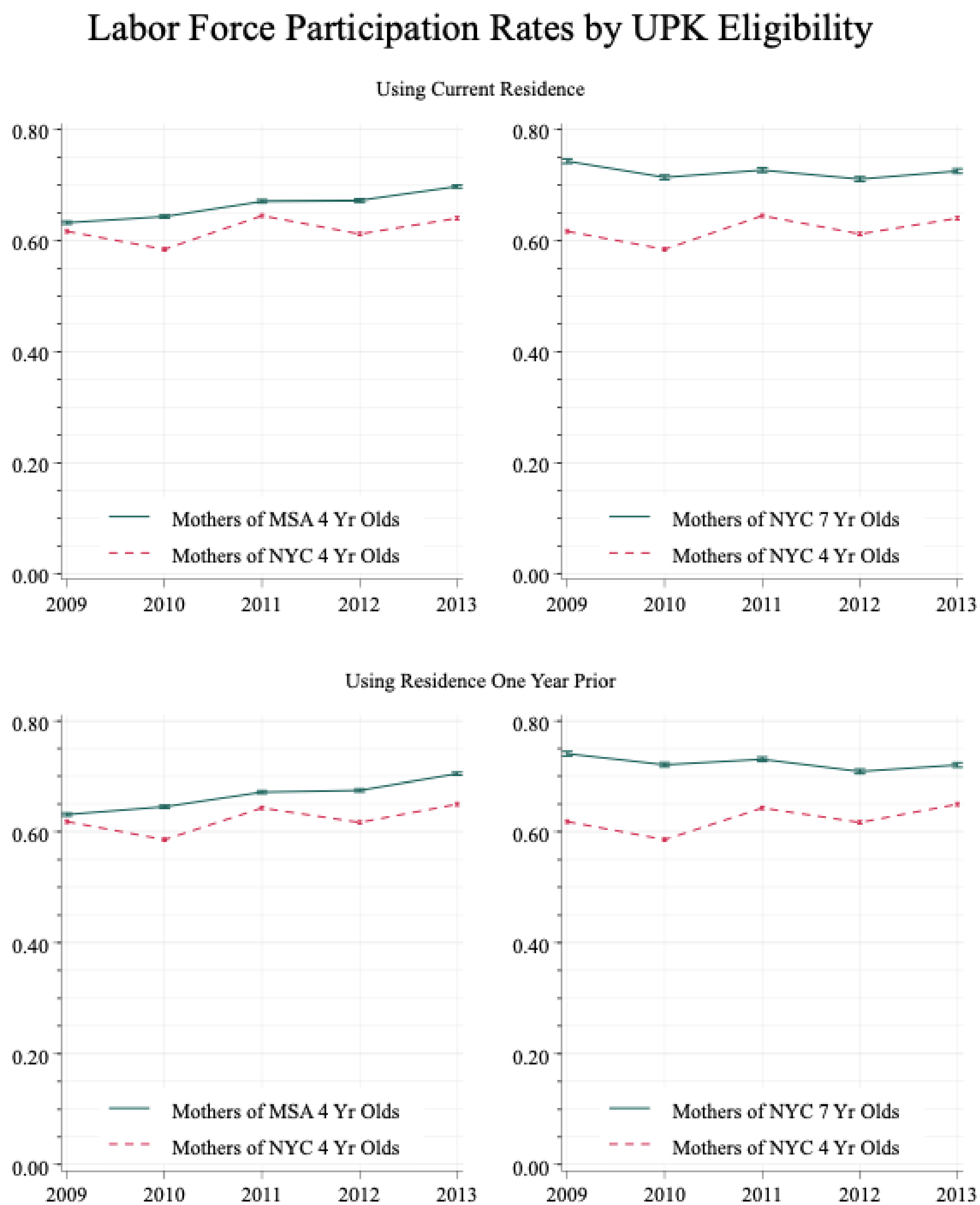
Table 19: Difference-in-Differences Results, Labor Force Participation (Migration Adjusted)

Comparison Group:	2010-2014		2010-2016	
	(1) No Cov.	(2) Covariates	(3) No Cov.	(4) Covariates
Mothers of MSA 4 Yr Olds				
Post Policy × Eligible	0.055	0.065*	0.028	0.035*
	(0.029)	(0.028)	(0.016)	(0.015)
Post Policy	−0.053*	−0.063**	−0.005	−0.010
	(0.024)	(0.024)	(0.020)	(0.019)
Eligible	−0.050**	−0.042*	−0.050**	−0.038*
	(0.017)	(0.017)	(0.017)	(0.017)
Constant	0.702***	0.122	0.702***	0.081
	(0.015)	(0.124)	(0.015)	(0.099)
R^2	0.004	0.077	0.004	0.077
Adjusted R^2	0.003	0.075	0.003	0.076
Observations	10,121	10,121	15,747	15,747

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Notes: The sample is selected using mothers' residence one year prior, rather than their residence in the current year.

Figure 2: Assessing the Bias Stability Assumption with Migration Adjustment: Mothers' Unconditional Labor Force Participation Rate by UPK Eligibility



Notes: Figure 2 plots mothers' pre policy (2009-2013) labor force participation rates by UPK eligibility, offering a comparison between using mothers' current residence (top panel) and mothers' residence one year prior (bottom panel) to determine sample inclusion and UPK eligibility. Each column corresponds to a different DID comparison group. NYC mothers in comparison groups are excluded if they have a child under age 6. Observations are weighted using ACS person weights.

Table 20: Difference-in-Differences Results, Labor Force Participation (Migration Adjusted)

Comparison Group:	2010-2014		2010-2016	
	(1)	(2)	(3)	(4)
	No Cov.	Covariates	No Cov.	Covariates
Mothers of NYC 7 Yr Olds				
Post Policy × Eligible	0.074**	0.073**	0.032	0.034*
	(0.025)	(0.023)	(0.017)	(0.017)
Post Policy	−0.062*	−0.062**	−0.001	−0.003
	(0.025)	(0.023)	(0.026)	(0.026)
Eligible	−0.096***	0.026	−0.096***	0.019
	(0.014)	(0.024)	(0.014)	(0.021)
Constant	0.739***	0.036	0.739***	0.033
	(0.020)	(0.152)	(0.020)	(0.106)
R^2	0.008	0.100	0.008	0.101
Adjusted R^2	0.007	0.097	0.008	0.098
Observations	6,760	6,760	10,675	10,675

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Notes: The sample is selected using mothers' residence one year prior, rather than their residence in the current year.

Table 21: Difference-in-Differences Results, Usual Hours Worked Per Week (Migration Adjusted)

Comparison Group:	2010-2014		2010-2016	
	(1) No Cov.	(2) Covariates	(3) No Cov.	(4) Covariates
Mothers of MSA 4 Yr Olds				
Post Policy × Eligible	2.550*	2.710*	0.381	0.519
	(1.118)	(1.066)	(0.689)	(0.666)
Post Policy	−2.240*	−2.555**	0.056	−0.327
	(0.966)	(0.940)	(0.835)	(0.824)
Eligible	−1.927**	−1.022	−1.927**	−0.823
	(0.715)	(0.627)	(0.714)	(0.618)
Constant	25.295***	0.802	25.295***	−4.152
	(0.643)	(4.732)	(0.643)	(3.662)
R^2	0.006	0.095	0.006	0.097
Adjusted R^2	0.005	0.093	0.005	0.096
Observations	10,122	10,122	15,749	15,749

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Notes: The sample is selected using mothers' residence one year prior, rather than their residence in the current year.

Table 22: Difference-in-Differences Results, Usual Hours Worked Per Week (Migration Adjusted)

Comparison Group:	2010-2014		2010-2016	
	(1) No Cov.	(2) Covariates	(3) No Cov.	(4) Covariates
Mothers of NYC 7 Yr Olds				
Post Policy × Eligible	2.777*	2.653*	1.857*	1.890*
	(1.267)	(1.165)	(0.837)	(0.785)
Post Policy	−2.248*	−2.232*	−1.631	−1.942
	(1.136)	(1.067)	(1.185)	(1.131)
Eligible	−3.471***	0.634	−3.471***	0.842
	(0.689)	(1.142)	(0.689)	(0.968)
Constant	26.620***	−2.404	26.620***	−5.346
	(0.835)	(5.617)	(0.835)	(3.848)
R^2	0.009	0.117	0.007	0.120
Adjusted R^2	0.008	0.115	0.007	0.118
Observations	6,761	6,761	10,676	10,676

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 23: DDD Results, Usual Hours Worked Per Week (Migration Adjusted)

Comparison Group:	2010-2014		2010-2016	
	(1) No Cov.	(2) Covariates	(3) No Cov.	(4) Covariates
Mothers of NYC 7 and MSA 4 and 7 Yr Olds				
Post Policy × 4 Yr Old=1 × Mom Lived in NYC Last Yr=1	6.393***	6.606***	2.792*	2.955**
	(1.913)	(1.830)	(1.159)	(1.114)
Post Policy × 4 Yr Old=1	−3.610*	−3.854**	−0.931	−0.993
	(1.447)	(1.401)	(0.791)	(0.788)
Post Policy × Mom Lived in NYC Last Yr=1	−3.843**	−3.915**	−2.414*	−2.444*
	(1.472)	(1.444)	(1.053)	(1.006)
4 Yr Old=1 × Mom Lived in NYC Last Yr=1	−2.488*	−2.234*	−2.488*	−2.275**
	(0.981)	(0.881)	(0.980)	(0.876)
Post Policy	1.629	1.574	0.996	0.680
	(1.051)	(1.067)	(0.815)	(0.858)
4 Yr Old=1	−0.990	2.356*	−0.990	2.299**
	(0.693)	(0.942)	(0.693)	(0.878)
Mom Lived in NYC Last Yr=1	0.561	1.129	0.561	1.361
	(0.920)	(0.895)	(0.920)	(0.882)
Constant	26.025***	−4.622	26.025***	−7.810*
	(0.738)	(4.377)	(0.738)	(3.310)
R^2	0.008	0.089	0.007	0.092
Adjusted R^2	0.007	0.088	0.006	0.090
Observations	13,951	13,951	21,829	21,829

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Notes: The sample is selected using mothers' residence one year prior, rather than their residence in the current year.