Spring 2015

# Three Essays in the Economics of Child Health and Development

Jia Gao
*University of New Hampshire, Durham*

# THREE ESSAYS IN THE ECONOMICS OF

# CHILD HEALTH AND DEVELOPMENT

BY

JIA GAO

Baccalaureate Degree (B.S.), Nankai University, 2009

Master's Degree (M.A.), University of New Hampshire, 2010

DISSERTATION

Submitted to the University of New Hampshire

in Partial Fulfillment of

the Requirements for the Degree of

Doctor of Philosophy

in

Economics

May, 2015

This dissertation has been examined and approved in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics by:

Dissertation Chair, Reagan A. Baughman, Associate Professor of Economics
Department of Economics, University of New Hampshire

Le Wang, Assistant Professor of Economics
Department of Economics, Finance, Legal Studies, the University of Alabama

Andrew James Houtenville, Associate Professor of Economics and Research
Director of the Institute on Disability
Department of Economics, University of New Hampshire

Robert D. Mohr, Associate Professor of Economics
Department of Economics, University of New Hampshire

Robert S. Woodward, Forrest D. McKerley Professor of Health Economics
Department of Economics, University of New Hampshire

On May 1st, 2015

Original approval signatures are on file with the University of New Hampshire Graduate School.

# DEDICATION

This dissertation is lovingly dedicated to my parents and many friends. Their support, encouragement, and constant love have sustained me throughout my life.

I would especially like to thank my Mom, Meiling Kang, who opened my eyes to the world and encouraged me to pursue my dreams, and my Dad, Guanlong Gao, who inspired my interest in economics and challenged me to further my efforts when I wanted to give up. I also dedicate this work to my many friends who have supported me throughout the process. I will always appreciate all they have done. I would like to give special thanks to my friend Xi Chen for being there for me throughout the entire doctoral program. All of you have been my best cheerleaders.

# ACKNOWLEDGEMENTS

I would like to take this opportunity to thank everyone who provided support and helped me throughout this process.

My deepest gratitude goes first and foremost to my dissertation advisor, Dr. Reagan Baughman, who has guided and supported me over the last five years. I appreciate so much her exquisite attention to detail and her demand for excellence. I would never have been able to finish my dissertation without her guidance, caring and patience. She has set an example of excellence as a researcher, mentor, teacher, and role model.

I also would like to express my heartfelt gratitude to one of my committee members, Dr. Le Wang, whose passion in research has always motivated me and whose insightful thoughts and innovative ideas have helped me greatly in my work. His enthusiasm and dedication to economics and education is truly inspiring and remarkable. Despite his busy schedule, Le has always found the time to talk with me—from exploring his innovative research ideas and the latest quantitative methods to sharing his experiences in teaching and publishing. He is one of the smartest, most knowledgeable and diligent scholars I have ever met.

I would like to express my earnest appreciation towards my other committee members—Dr. Andrew Houtenville for helping me to apply for the restricted dataset, providing me a collaborative funding opportunity over the summer, and giving me thoughtful suggestions; Dr. Robert Mohr for using his expertise to help me develop the theoretical model; and Dr. Robert Woodward for his countless hours of reading and reflecting on my work, and most of all for his patience throughout the entire process.

I would like to extend my thanks to the UNH Economics Department for allowing me to conduct my research. My special thanks go to the staff of my department, Sinthy Kounlasa and Stacy Hokinson, for their excellent administrative support and hard work. I wish to express my gratitude to the Institutional Review Board for the Protection of Human Subjects in Research (IRB) at UNH for approving the project in this dissertation. I also would like to acknowledge the UNH Graduate School and the UNH Economics Department for providing me with the Dissertation Year Fellowship to support this work.

I am grateful to many people who shared their memories and experiences in writing their own dissertations, especially Minghua Li who was always willing to help and give her best suggestions. I must acknowledge as well the many friends, classmates, teachers, and colleagues who assisted, advised, and supported me in my research and writing efforts over the years. Especially, I need to express my gratitude and deep appreciation to Xi Chen, Zhao Pan and Yunjin Sun whose companionship, knowledge, and wisdom have supported, enlightened, and entertained me over the years of our friendship. They have consistently helped me keep perspective on what is important in life and have shown me how to appreciate each day. My sincere thanks also go to my classmates and friends at UNH who helped me in many different ways and made this journey more memorable—Jennifer Trudeau, Son Nguyen, Michael Kurtz, Wei Shi and Chao Liu.

Most importantly, I would like to thank my parents for their love, support, understanding and sacrifices. They were always there cheering me up and stood by me through the good times and bad.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# ABSTRACT


THREE ESSAYS IN THE ECONOMICS OF

CHILD HEALTH AND DEVELOPMENT

by

Jia Gao

University of New Hampshire, May, 2015

This dissertation is composed of three independent essays that focus on the economics of child health and development.

The first chapter explores whether availability of the SBP has affected maternal labor supply by using variation in the SBP mandates within-state over time to identify the effect. To increase the availability of the School Breakfast Program (SBP), between 1989 and 2012, 21 states passed laws that require schools to provide the SBP if the fraction of students eligible for free or reduced-price breakfast in their school districts exceeds a certain threshold. Using the CPS Food Security Supplement data between 1995 and 2012, I first show that the SBP mandates significantly increase program participation among mothers with a high school degree or below and among single mothers. Then I estimate the effects of mandates on maternal labor supply using March CPS 1990 to 2013 surveys. The findings suggest that among less-educated mothers and single mothers, a mandate that requires all schools to provide the SBP is associated with an increase in the probability of being employed and working full time, and an increase in weekly hours of work. However, weaker mandates do not have the same maternal labor supply effects.

The second chapter examines the effects of smoking bans on birth outcomes. Prenatal smoking has serious adverse consequences on infant health. Among the newest policies developed to reduce smoking and second-hand smoke are smoking bans. Using individual-level birth certificate data from the Natality Detail File between 1995 and 2009, which is matched to county-level data on smoking bans, we investigate the impacts of smoking bans in bars, restaurants and workplaces on infant birth weight, gestations, 5-minute APGAR scores and incidences of cleft lip/palate. In general, bans are not associated with changes in birth weight or weeks of gestation. Surprisingly, we find small increases in rates of low birth weight and very low birth weight infants born to young women in counties that adopted at least one type of ban during the study period. We also show that the negative infant health effects associated with smoking bans appear among babies born to mothers who reported not smoking during pregnancy; this suggests that increased exposure to second-hand smoke is likely to be the mechanism.

The third chapter examines the impact of a heretofore relatively unexplored input in the educational process—language environment at home—on student academic achievement during early childhood. Using the confidential data from Early Childhood Longitudinal Study-Kindergarten Class (1998-99), we are able to exploit cross-sectional geographic variation in local language environment, augmented with the recently developed instrumental variable strategy in Lewbel (2012), to identify the causal effect. Our results show that speaking a language other than English at home has a sizable, negative impact on reading test scores in both third and fifth grades, but has no effect on math scores in either grade. We find no evidence that speaking a language other than English at home has any effect on the growth rate of test scores from the third to the fifth grade, regardless of the subject.

# INTRODUCTION

Among OECD countries, United States ranks relatively low on many measures of child well-being. For example, the U.S. infant mortality rate ranks 27$^{th}$ among OECD countries and has nearly double the rate of countries like Japan. Child well-being, measured by health status, school performance, and cognitive and non-cognitive skills, is an important predictor of future health, education, and labor market outcomes. Therefore, an analysis of the determinants of child well-being is not only important to the welfare of children but also to inform the development of a variety of public policies. This dissertation evaluates the effects of three specific determinants of child well-being.

The first chapter is a study of the effects of a federally sponsored nutrition program—the School Breakfast Program (SBP)—on maternal labor supply. Established as a pilot program in 1966, the SBP provides free or reduced-price breakfast to children from low-income families with the goal of improving their nutritional intake. In addition to the health and nutrition effects of this programs on children, the SBP may affect the behaviors of other household members due to a redistribution of resources within the family. For instance, the program may create incentives for mothers to change their labor supply by creating notches in the budget constraint and reducing time spent on household production. This study is the first to estimate the labor supply impacts of the SBP. The results show that among mothers with high school education or less and single mothers, the strongest form of mandate, which requires all school districts to provide breakfast, is associated with statistically significant increases in the probability of being employed and in number of hours of work. Given that the labor supply effect of means-tested transfer programs is a central concern for policy makers this research sheds light on the design of a means-tested transfer program that provides a valuable benefit to recipients and may help them to attain self-sufficiency at the same time.

The next chapter analyzes the impacts of clean indoor air laws on infant health. One of the leading causes of infant mortality and other negative birth outcomes in the U.S. is maternal tobacco use. As a more recently developed anti-smoking policy, smoking bans are designed to decrease smoking rate and reduce exposure to second-hand smoke. The literature has looked at the impact of smoking bans on smoking behaviors, but the effect of smoking bans on exposure to second-hand smoke is not clear. Further, how smoking bans affect the public health through the channel of second-hand smoke is not well understood. Thus, the second chapter investigates the impact of smoking bans on several infant birth outcomes. The results show that smoking bans are not associated with any increase in birth weight or weeks of gestation. This suggests that, unlike cigarette taxes, smoking bans are not an effective policy instrument for addressing infant health problems caused by smoking, and may actually result in adverse effects.

Another important factor in children's early childhood development is the language that they speak. Language environment at home is one of the most important factors in explaining the test score gap between children of immigrants and those of natives. The third chapter estimates the effect of speaking a language other than English at home on academic achievement during early childhood. The results show that the use of a language other than English at home decreases student reading scores in both third and fifth grades, while there is no significant effect on math scores. There is no evidence that language environment at home is correlated with the growth rates in the test scores, regardless of subject. Since cognitive and non-cognitive skills developed in early childhood can pre-determine one's future educational and market outcomes, our results raise serious concerns about the long-term well-being of the children of immigrants whose primary language at home is not English.

# CHAPTER 1

## THE EFFECT OF THE SCHOOL BREAKFAST PROGRAM ON MATERNAL LABOR SUPPLY[±]

Jia Gao

Department of Economics

University of New Hampshire

[±] Acknowledgements: I would like to thank my advisor Reagan Baughman, and my committee members, Le Wang, Andrew Houtenville, Robert Mohr and Robert Woodward for their comments and suggestions. I am grateful to Kitt Carpenter, Christina Robinson, and the participants of UNH Economics seminar for helpful comments.

## 1. Introduction

The School Breakfast Program (SBP) served more than 12 million children per day in both public and non-profit private schools at a cost of $3 billion to the federal government in the 2012-2013 school year.[1] Established as a federal pilot program in 1966, the SBP became a permanent program in 1975. The initial goal of this program was to provide subsidized breakfast to students from low-income families who might not be able to afford a nutritious meal. According to the eligibility rules of the SBP, children from families with incomes at or below 130 percent of the Federal Poverty Line (FPL) can eat school breakfast for free, and children with household incomes between 130 and 185 percent of the FPL are eligible for reduced-price breakfast.[2] Students with family incomes above 185 percent of the FPL can still participate in the program but they must pay full price. Regardless of income, for many working parents whose opportunity cost of preparing a nutritious breakfast is high, the availability of this program, even at full price, makes it possible for students who live in busy families to eat breakfast at school.

The research literature has documented the impact of the SBP on child nutrition and health, and even on school performance, but almost all of these studies examine the SBP effects only on the immediate recipients: the students who receive the SBP. The program may have also affected the behavior of other family members, but very few studies have looked at this question.[3] Children are part of an economic unit within a family, and the SBP that targets children may create

---

[1] The SBP statistics are available at: http://www.fns.usda.gov/sites/default/files/SBPfactsheet.pdf, accessed September 2013.

[2] In 2014, 130 percent of the FPL is $31,005 for a household of four, and 185 percent of the FPL is $44,122 for the same household.

[3] Bhattacharya et al. (2006) estimate the SBP effect on nutrition outcomes of other household members. They find that the SBP improves two measures of dietary quality for other household members, but has no significant effects on nutrition outcomes. They argue that this lack of significance is due to small sample size.

incentives for other family members. For example, the food provided for children might lead to a redistribution of resources within a family; parents may feed participating children less at dinner and feed non-participating children more. Additionally, the SBP may affect parents' labor supply by creating notches in the budget constraint and reducing the time spent on household production. As we know, the labor supply effect of means-tested transfer programs is a concern for policy makers. Therefore, this paper provides an analysis of the work incentive effects of the SBP on mothers.

There are two mechanisms by which the School Breakfast Program could affect maternal labor supply. First, the SBP may create the same work disincentives as the other means-tested transfer programs because it provides free or reduced-price breakfast only to children from low-income families. These families are inframarginal: they treat one dollar worth of food benefit the same as one dollar in cash income (Hoynes and Schanzenbach 2009). As a result, the subsidized breakfast offered by the SBP can be considered an income transfer, and this generates employment disincentives. Given that the income eligibility threshold is 130 percent of the FPL for free breakfast and 185 for reduced-price breakfast, mothers living in households with incomes close to 130 (185) percent of the FPL may find it in their best interest to reduce labor supply so as to qualify for the subsidized breakfast as long as the value of the breakfast is greater than the foregone wages.

The magnitude of these eligibility effects on labor supply depends on the monthly dollar value of the SBP. The more school-aged children a family has, the larger the work disincentives. For a female headed household with three children eating free school breakfasts, assuming the dollars' worth of a breakfast is equal to the federal reimbursement rate for a free breakfast ($1.58

in 2013),[4] the noncash monthly transfer is roughly equal to $95 (1.58*3*20). This amount is approximately equal to earnings from 13 hours of work at the federal minimum wage rate. However, the implicit market price of school meals is likely to be higher than the reimbursement rate because: (1) schools must operate the SBP on a non-profit basis, (2) 64 percent of schools serve breakfasts that cost more to prepare than the reimbursement rate (Bartlett et al. 2008), and (3) there exists increasing economies of scale; compared with individual households, schools can purchase foods at a lower price due to greater purchasing power stemming from the schools' size (Ollinger et al. 2011). Considering all these factors, it is very likely that a representative household values a month of school breakfasts at more than $95.[5]

Second, independent of the effect of the in-kind transfer, availability of the SBP itself may increase maternal employment by reducing the time spent on household production. Mothers often face tradeoffs between employment and household production (see, for example, Apps and Rees 1996). By "outsourcing" children's meals, mothers can reallocate time and energy to paid work, other types of household production, or even leisure. One descriptive study shows that the usual meal preparers whose children obtain meals at school spend 3.7 more hours per week in paid work, and 1.3 fewer hours in child care, than those whose children do not eat school meals (Hamrick et al. 2011). In this sense, the SBP creates the incentive for mothers to work longer hours, regardless of whether their children eat the breakfast provided by this program for free or at full price.[6]

---

[4] In the 2013-2014 school year, the federal cash reimbursement rates for breakfasts are $1.58 for free breakfasts, $1.27 for reduced-price breakfasts and $0.29 for full-price breakfasts.

[5] Compared with the Supplemental Nutrition Assistance Program (SNAP), which provided an average monthly benefit of $275 per household in 2013, the SBP is a relatively smaller transfer. Currently a family of four has a maximum guarantee of $649 per month in SNAP, and this benefit is phased out using a benefit reduction of 30 percent.

[6] On average, mothers aged 18-64 spent 39 hours on household work (including child care and shopping) per week during 2003-2008, almost double the time spent by fathers (Bianchi 2011). Since mothers are the primary meal

In this paper, I employ a quasi-experimental approach to study the effect of the SBP on maternal labor supply. To increase the availability of the SBP, many states mandate that school districts provide the SBP if the fraction of students eligible for free or reduced-price breakfast exceeds a specific threshold. In 1976, South Carolina passed legislation requiring school districts to provide the SBP if 40 percent or more of students were eligible for subsidized breakfast. This was the first state mandate. By 1988, another four states had implemented mandates to require certain school districts to offer the SBP. Between 1989 and 2012, 21 additional states enacted some type of SBP mandate, and the percentage of mothers living in a state with the SBP mandate increased from 14.74 to 65.29 percent. The variation in the timing of state mandate adoption is used to identify the impact of SBP availability on maternal labor supply. Using the CPS Food Security Data from 1995 to 2012, I first show that state mandates have a strong effect on school breakfast participation. Then this paper investigates the work incentive effects of the SBP mandates using March CPS data between 1989 and 2012. In the full sample of all mothers, I do not find any significant effects of mandates on maternal employment. But among mothers with a high school diploma or less and single mothers with any level of education, the results indicate that strong mandates that require all school districts to provide the SBP encourage labor supply. Specifically, strong mandates increase the probability of being employed by 2.6 percentage points, the probability of working full time by 3.4 percentage points, and the hours of work by 1.25 hours per week for less-educated mothers. The effect of strong mandates is even higher among single mothers. Weaker mandates that only ask certain school districts to participate in the SBP do not have the same labor supply effects.

---

preparers and care givers for children, I assume the effect of the SBP on labor supply is felt mainly by mothers rather than fathers.

This paper contributes to the literature in two ways. First, although previous research has analyzed labor supply effects of many means-tested transfer programs, this is the only study that estimates the impact of the SBP on maternal labor supply. The findings of this study underscore the importance of evaluating the indirect effects of nutrition programs when designing transfer programs. Second, since the SBP is a federal program, it is difficult to use a quasi-experimental approach to evaluate the effects of this program due to lack of state variation. This is one of the very few studies that uses the variation in state mandates about school participation to evaluate the effects of the SBP.[7]

The paper proceeds as follows. Section 2 reviews past research on the labor supply effects of means-tested transfer programs and literature on household production and maternal employment. Section 3 outlines a conceptual framework for the effect of the SBP on labor supply. Section 4 provides information about state mandates. Section 5 and 6 lay out the empirical strategy and data used in this paper. Results and conclusions are discussed in Sections 7 and 8, respectively.

## 2. Related Studies

Although there is a vast literature on the labor supply effects of transfer programs, no existing studies have examined the work incentives generated by either the School Breakfast Program or the National School Lunch Program. Given that the SBP could affect maternal employment either by creating notches in the budget constraint or by reducing the amount of time needed for household production, it is useful to review two strands of the literatures: (1) studies of

---

[7] To the best of my knowledge, only Frisvold (2012) uses the variation in the SBP mandates' thresholds across states in the year of 2004 to estimate the effect of the SBP on student cognitive development.

other means-tested transfer programs and female employment, and (2) research on tradeoff between household production and labor supply.

## 2.1 *Means-tested Transfer Programs and Work Incentives*

Existing research provides evidence that many means-tested transfer programs discourage labor force participation and reduce weekly hours of work (for reviews, see Danziger et al. 1981; Moffitt 1992; Hoynes 1997; Currie 2003; Moffitt 2002). Most of these studies examine the work disincentives for female headed households, the largest recipient group in many transfer programs. Of all cash and noncash transfer programs, the Supplemental Nutrition Assistance Program (SNAP, formerly called the Food Stamp Program, or FSP) is most similar to the SBP because both are federally funded food and nutrition programs and provide in-kind transfers to households. Therefore, I mainly focus on the research on the FSP/SNAP. Aside from SNAP/FSP, it is also important to review the literature on Temporary Assistance for Needy Families (TANF, or Aid to Families with Dependent Children, AFDC before 1996) because a considerable fraction of SBP participants are TANF/AFDC recipients and the labor supply effect of AFDC is well studied. In theory, TANF/AFDC and SNAP/FSP both discourage labor supply for two reasons. First, the existence of a guaranteed benefit amount without any work/earnings creates work disincentives. Women may reduce hours worked or not take a job in order to become income-eligible for the benefit. Second, the benefit reduction rate (or implicit tax rate) reduces the incentives to work because the benefit decreases as the earnings increase. This benefit reduction rate is 30 percent in the SNAP/FSP, and now varies by state for TANF (ranging from 0 to 100 percent).[8]

[8] The benefit reduction rate was 100 percent in AFDC from 1981 to 1996.

The literature on the labor supply effects of the Food Stamp Program (FSP) is relatively small, and there are no studies that look at SNAP. Fraker and Moffitt (1988) is the first study that examines the impact of FSP participation on labor supply for female headed households. The authors adopt a structural approach and use maximum likelihood to jointly estimate female labor supply and program participation. They find that FSP participation reduces work by one hour per week or 9 percent. Given that married couples may determine their labor supply simultaneously, Hagstrom (1996) uses a nested multinomial logit model to examine the intrafamily labor supply impact of the FSP. Both his structural model estimates and simulation results indicate weak labor supply effects for husbands and wives. He concludes that the labor supply of a married couple shows little response to changes in the guaranteed benefits and benefit reduction rate of the FSP.

Since there is substantial overlap in participation between the FSP and other means-kind transfer programs, some studies estimate the work disincentives of multiple transfer programs simultaneously. For example, as an extension of Fraker and Moffitt (1988), Keane and Moffitt (1998) study the labor supply impact of several transfer programs, including the FSP. Their simulation results indicate that small to moderate reductions in the benefit reduction rate increase weekly hours of work for existing FSP recipients but lead to more recipients at the same time. Axelsen et al. (2007) investigate how welfare recipients make their employment decision based on transfer benefits and other factors. The authors conclude that transfer benefits affect not only the decision of whether or not to work, but also the decision of how many hours to work. Their results suggest that increased benefits in the FSP discourage work effort. Huffman and Jensen (2005) show that FSP participation and TANF participation are highly correlated, and participation in these two programs jointly affects the work decision.[9] Notably, all of these studies employ structural models and use only cross-sectional variation in FSP benefits, which vary by family size.

Researchers use structural models to assess labor supply effects of the FSP because there is little cross-state or over-time program variation in eligibility rules or benefits to exploit, but their results are sensitive to the specification of the utility functions and stochastic assumptions (for a review, see Moffitt 2002). To avoid these problems, Hoynes and Schanzenbach (2012) use a quasi-experimental approach to study how the historical availability of the FSP affected labor supply and earnings. Using variation in the timing of county-level introduction of the FSP during 1960s and 1970s, they find that for a female headed household, the introduction of the FSP reduced women's employment by 183 hours per year on average, which translates to a reduction of 505 hours among FSP participants. For the overall population, the authors do not find any significant effect of the FSP on labor supply and attribute the insignificance to a low FSP participation rate.

Compared to the studies of the FSP, the literature on labor supply effects of TANF/AFDC is sizable. Historically, researchers found that AFDC had a greater negative effect on labor supply than did other transfer programs, and the effect was bigger for single women than married couples (Hoynes 1997). Early studies (for example, Hausman 1980; Moffitt 1983; Hoynes 1996) primarily adopt a structural approach by using cross-state variation, and show that the AFDC participation reduces female labor supply by 10 to 50 percent. Meyer and Rosenbaum (2001) also employ a structural approach to investigate the work incentive effect of AFDC and find that a one thousand dollar reduction in annual welfare benefits increases employment among single mothers by 3 percentage points. Later studies, particularly studies after 1996, use reduced-form models and

[9] Some sociological studies, such as Livermore and Powers (2006) and Elliott and Packham (1998), focusing on the role that transfer programs play in welfare-to-work decisions, will not be reviewed here due to the space limitation.

employ within-state over-time variation to evaluate AFDC waivers and TANF program (for example, Moffitt 1999; Schoeni Blank 2000). A majority of studies find that the existence of AFDC waivers and the implementation of TANF increase female employment due to the time limit on welfare participation and the termination of benefits under a work requirement.

### 2.2 *Maternal Labor Supply, Household Production and Child Care Costs*

As discussed earlier, the availability of the SBP, either at full price or reduced price, may affect a mother's time allocation between household production and hours of paid work. To understand the link between maternal labor supply and household production, it is useful to review the literature on the theory of allocation of time and household production. Becker (1965) is the first to propose a theory of household production, where individuals maximize utility subject to time and budget constraints. Utility is a function of commodities that are produced using market goods and individuals' time. Although Becker introduces consumer's time, along with goods and services, into a consumer production function, his theory does not separate leisure from the consumer's production time. To differentiate housework and leisure, Wales and Woodland (1977) formulate a work-leisure-household production model in which income, leisure and household production are determined simultaneously. This is the first model to treat household production as an endogenous variable, equal to the total time net of work and leisure. Gronau (1977) provides further justification for establishing the distinction between housework and leisure in the theory of allocation of time. This distinction, he argues, is a prerequisite for any study of time allocation and is extremely useful in the analysis of fertility, gains from marriage, demand for child care as well as labor force participation. More recently, Apps and Rees (1996) demonstrate the importance of incorporating household production in the labor supply model theoretically and empirically.

Cherchye et al (2012) provide empirical evidence that adults' preferences are not only affected by leisure and the consumption of market goods, but are also dependent on the consumption of home produced goods.

Although there is no empirical work that analyzes how preparing meals for children affects maternal employment, there exist numerous studies on another home produced item: child care. The cost of child care can affect a mother's choice of occupation, place of employment, and hours of work (both the length and the start and end time). Researchers have examined child care availability, price and cost on the labor market behaviors of women with children, and most found low child care costs/prices/expenditures, and high child care subsidies increase maternal labor supply. The estimated child care price elasticities for maternal labor supply vary from substantially negative (-1.26) to close to zero (for reviews, see Anderson and Levine 2000; Blau and Tekin 2001; Blau 2003).[10]

The empirical studies on child care costs and maternal labor supply can be classified into three categories. The first category investigates the imputed or predicted child care price effects on labor supply, using either a reduced-form model or a structural approach (Ribar 1995; Kimmel 1995; Baum 2002; Connelly and Kimmel 2003). Studies within this category suggest a negative relationship between child care prices and maternal employment and show that employment elasticities are larger for single than for married mothers. The second group is reduced-form studies that estimate the impact of actual child care subsidies on maternal employment (Berger and Black 1992; Averett et al. 1997; Tekin 2005; Tekin 2007; Blau and Tekin 2007). Although, theoretically, non-linear child care subsidies generate work disincentives for certain mothers, most empirical

[10] Different model specifications and econometric methodologies are mainly attributed to this large variation in results, though differences in data resources and sample composition (by child age and marital status) also account for the variation (Blau 2003).

research indicates that subsidies increase maternal labor supply overall. In particular, they help single mothers to gain economic self-sufficiency through work. The last cluster of studies use a quasi-experimental approach to assess the effect of child care subsidies on labor supply. Using variation from exogenous changes in child care subsidies, these studies tend to generate larger negative price elasticities (Gelbach 2002; Lefebvre and Merrigan 2008; Baker et al. 2008; Herbst 2010).

## 3. Theoretical Framework

Following Gronau (1977), I build a labor supply model for a single mother in order to study the impact of the SBP on maternal employment. To simplify the model, the situation in which a husband and wife make joint decisions in the labor market is not considered. Assume a mother has a quasi-concave and increasing, twice-differentiable utility: $U(C, L)$, where $C$ is total household food consumption, including breakfast, and $L$ is pure leisure for herself.[11] Suppose $C$ and $L$ are normal goods. The mother's consumption is composed of two types of goods ($C = X + H$): market good, $X$, with price unity; and home produced good, $H$, determined by the time spent on household production. Assume these two goods are perfect substitutes. The mother's time ($T$) is spent on household production ($t_h$), market work ($t_w$), and leisure ($L$), and $T = t_h + t_w + L$. The household production function $H = h(t_h)$ is strictly increasing, twice-differentiable, and concave.[12] Formally, I build the following model:

[11] This model can easily generalize to all other goods as long as the other goods are separable from $U$. Let $Y$ denote non-food consumption. The new utility function $V(Y, U(C, L))$ will produce exactly the same Equations (3) and (4) at the equilibrium.

[12] For simplicity, time is the only input into household production.

$$\max_{X,t_h,L} U = U(X + h(t_h), L) \tag{1}$$

$$s.t. (T - L - t_h) * W + a = (1 - s) * X \tag{2}$$

where $a$ is fixed non-labor income; $W$ is wage rate; $s$ is the implicit subsidy of the SBP, which is defined as the difference between the reimbursement rate of the SBP and the market value of the SBP.

If the mother works in the market, there is a solution for an interior optimum by maximizing the Lagrangian function with respect to $X$, $t_h$ and $L$. The following equations result from the first order conditions at the equilibrium:

$$\frac{\frac{\partial U}{\partial L}}{\frac{\partial U}{\partial C}} = \frac{W}{1-s} \tag{3}$$

$$\frac{W}{1-s} = h'(t_h) \tag{4}$$

Equation (3) suggests that the marginal rate of substitution between consumption and leisure is equal to the shadow price of time, $W/(1-s)$, or the real wage rate. By assuming home goods and market goods are perfect substitutes, I obtain Equation (4), which indicates that the marginal product of household production is equal to the shadow price of time. In other words, one unit of time can buy one unit of goods. If the mother does not work in the market, then:

$$h'(t_h) = \frac{\frac{\partial U}{\partial L}}{\frac{\partial U}{\partial C}} = W^* > \frac{W}{1-s} \tag{5}$$

where $W^*$ is the reservation wage, and $\frac{W}{1-s}$ is the real wage. In this case, the mother is self-employed at home because her marginal productivity at home is greater than the wage rate she would be paid in the market.[13]

[13] Another scenario is that the mother only works in the market but not at home. If the marginal productivity of household production is always less than the real wage rate, there is no household production. Then I turn to the familiar dichotomy of work and leisure.

This model is graphically depicted in Figure 1. A mother, whose preferences are denoted by indifference curve $U$, chooses between consumption and leisure and is bound to the constraint line $EDFT$. The portion $ED$ is a straight line with slope $\frac{w}{1-s}$; the portion $DF$ is a concave curve, which describes the household production function; and the portion $FT$ denotes the non-labor income $a$. This constraint describes the following scenario: a mother buys the first item using non-labor income, produces additional goods at home until the marginal productivity of household production is equal to real wage rate, and buys the remaining goods from the market using her wage. If a mother participates in the labor force, she faces a tradeoff between working in the market and producing at home. The more time a mother spends on household production, measured by the horizontal distance from point $T$, the greater her consumption of home produced goods, measured by the vertical distance from the origin $O$. At the point $B$ she maximizes her utility, where she enjoys $L$ units of leisure and $C$ units of total consumption. The point $D$ is defined by Equation (4), where the opportunity cost of producing one more unit of home made goods is equal to real wage rate. For a mother who does not work in the market, her utility is also maximized when the indifference curve is tangent to the constraint. However, her point of tangency would be on the concave portion of the curve. At the optimal point, this mother only chooses between home goods and leisure.[14]

There are two mechanisms by which the SBP could affect maternal employment: the effect of the in-kind transfer and the effect generated by reducing household production. For simplicity, these mechanisms are modeled separately. I first focus on the effect of the in-kind transfers provided by the SBP by letting $s = 0$. Since this program provides free (reduced-price) breakfast

[14] If a mother only works in the market but not at home (which indicates that the marginal productivity at the point $F$ falls short of the real wage rate), the constraint will appear as a straight line with a slope equal to that of wage rate; as in the traditional Robbins diagram.

to children from families with income less than 130 (185) percent of the Federal Poverty Line, such in-kind transfers create two kinks in the budget constraint. For simplicity, I only draw one kink (185 percent of the FPL) in the graph. Assuming the amount of transfer is $X_b$, represented by the vertical segment in Figure 1, the new constraint is the curve $EMB'D'NT$. Only part of the budget constraint shifts outward in a parallel way because the transfer disappears when income is above 185 percent of the FPL. Under this new budget constraint, the optimal point could be a corner solution at $B'$, which shows the scenario where a mother decreases her labor supply so that her children remain eligible for the reduced-price breakfast.

The shift of the budget constraint from in-kind transfer affects neither the marginal product of household production (the shape of the curve *DF* does not change), nor the real wage rate. As a result, Equation (4) still holds and is represented by $D'$, indicating that the mother still spends $t_h$ on household production. Since both home goods and market goods are normal goods (including the SBP) by our assumption, the mother can attain higher utility by spending more time on leisure and less time on market work. Another scenario is that the mother whose children were originally eligible for the reduced-price breakfast may also decrease her labor supply because of the in-kind transfer. In this case, the new optimal point could be a tangent solution on line $B'D'$. In both scenarios, the effect of the in-kind transfer will unambiguously discourage maternal employment.

Next, I look at the second mechanism for how the availability of the SBP might affect maternal employment by assuming that all students pay the breakfast at full price and that the SBP could only affect mothers' time allocation. The availability of school breakfast (a market good) makes it much easier to substitute market goods for home goods. Namely, it is cheaper to buy the goods in the market rather than produce them at home. In the model, *s* denotes the implicit subsidy provided by the school breakfast, and $0 < s < 1$. According to Equation (4), the optimal time spent on household production $t_h^*$ decreases when $s$ increases from zero (here, the household production function has decreasing marginal product, that is, $h''(t_h) < 0$). As shown in Figure 2, there is a new budget constraint $KB'D'FT$ with a larger real wage rate $\frac{W}{1-s}$. The time for household production unambiguously declines. The consumption of home goods decreases, while the total consumption increases because both goods are normal goods.

The change in leisure in the second mechanism is ambiguous, and so is the change in market work. Three factors could affect leisure: (1) the reduction of household production could increase leisure; (2) the income effect incurred by the rise in real wage rate increases leisure; and (3) the substitution effect from the rise in real wage rate decreases leisure. If the marginal rate of substitution between consumption and leisure is large or the income elasticity of leisure is small, then leisure will decrease, and time spent on market work will increase. Even if the mother spends more time on leisure, her time spent on market work could still increase as long as the decline of time for household production is large. In other words, if the mother finances the additional consumption mainly through the wages, the time on market work will increase, but if the implicit subsidy provided by the SBP is enough for financing the additional consumption, then the mother may work less in the market. Thus, the effect of availability of the SBP on maternal employment is theoretically undetermined.

Based on this theoretical analysis, the direction of the maternal labor supply effect is ambiguous. The program generates both incentives and disincentives for mothers to work. Empirically, I test the net effect of this program to determine which mechanism dominates.

## 4. State SBP Mandates

Although the SBP is a federal entitlement program, students are only able to participate if the schools they attend offer the program. During the 1990-1991 school year, only four million children participated in the SBP, and only 48 percent of the schools offering the National School Lunch Program also provided the SBP. To increase availability of the SBP, Congress amended the Child Nutrition Act in 1989 by allocating start-up funds to state educational agencies for distribution to schools to offer and expand the SBP. These grants, which are different from the federal reimbursement for each meal, were used to cover the fixed costs associated with offering the breakfast. In practice, most of the grants were given to schools with a high percentage of low-income students and that promised to participate in the program for at least three years. This federal assistance was funded at a level of $5 million in fiscal year 1989, decreased to $3 million in fiscal year 1990, and ended in fiscal year 1996 (Fox et al. 2004).

Additionally, in order to encourage schools to accept start-up funds from the federal government, states began to require that some or all schools implement the SBP during the 1990s. The structure of most state laws is that schools must provide the SBP if the fraction of students eligible for free or reduced-price meals (i.e., whose family income is less than or equal to 185 percent of the FPL) exceeds a certain threshold. This state mandate is usually implemented at a school district level. The lower the threshold, the greater the number of schools that must participate in the program and the more likely a given student will gain access to the program. For example, New Jersey has a threshold of 0.2, which means that school districts with 20 percent or more students eligible for free or reduced-price meals have to provide the SBP. The threshold of zero means that all school districts must offer the SBP. Due to these state mandates, the size of the SBP more than doubled, from 3.8 million breakfasts per day in fiscal year 1989 to 8.1 million breakfasts per day in fiscal year 2002 (Fox et al. 2004). In the 1999-2000 school year, 75 percent of schools that provided the National School Lunch Program also served breakfast, and in the 2011-2012 school year this percentage went up to 90 percent (FRAC 2013).

Information about state mandates over the 1976 to 2013 period is collected from state statutes and documented in Appendix Table 1 in detail. Although the first state mandate dates back to 1976, when South Carolina required that school districts with 40 percent or more of students eligible for free or reduced-price meals must offer the SBP, most states began to enact mandates after the amendment of the Child Nutrition Act of 1989. Before 1990, state mandates were comparatively rare; only five states legislated mandates. But from 1990 to 1996, 14 more states implemented the mandates to require certain school districts to offer the SBP. Since 1996, another eight states have passed mandate laws. Currently, 27 states have mandate laws, and 28 states provide some type of funding for school breakfast. While the threshold of the mandate varies by state, most states have not changed their thresholds over time. Seven states, however, did lower their thresholds gradually over the first two or three years of legislation, so as to give schools more time to prepare for the implementation of the SBP. Among the states that have the mandate laws, eight states require all public schools to provide the SBP, and other states have thresholds below 0.4 with only two exceptions: Connecticut (0.8) and New Mexico (0.85).

## 5. Empirical Model

In many studies on the health and nutrition impacts of the SBP, researchers evaluate the effects of the program either by comparing participants with non-participants in the same school, or comparing students in schools that offer breakfast with students in schools that do not provide breakfast. However, this method may lead to selection bias because children are not randomly assigned into treatment. Some unobserved characteristics, such as mothers' preference about how much time is devoted to household production, may influence the SBP participation decision and the maternal employment decision simultaneously. Similarly, a school's decision to offer the SBP is not random. School-level (or district-level) unobservables like local residents' attitudes toward child nutrition and health could be associated with both maternal employment rate and the SBP participation rate. Additionally, there would be a reverse causality problem if a full-time employed mother were more likely to send her children to a school that offered the SBP.

Without an appropriate instrument for program participation, this endogeneity issue cannot be addressed. More recently, a reduced-form model has become the most common way to estimate the effects of programs, with a reliance on using variation in policy changes. A central challenge for applying this method to evaluate the SBP is that this program exhibits no variation across states or within each state over time in income eligibility rules or level of reimbursement for meals. However, as discussed in the previous section, each state implemented the SBP mandates at different times. Thus, I use the changes in the SBP mandates within each state over time as a source of variation to identify the effect of the SBP on maternal labor supply in this study.

Figure 3 illustrates the substantial growth in the percentage of mothers living in a state with the SBP mandate over the study period 1989 to 2012. As shown, the percentage of mothers living in a state with a certain type of mandate was only about 14 percent in 1989, but it went up rapidly to 50 percent in 1996, and finally reached 65 percent in 2012. Additionally, there was a moderate increase in the percentage of mothers living in a state with a full coverage mandate during this period. Here, the full coverage mandates are defined as the type of mandates that require all school districts to participate in the SBP. The ratio rose from less than one percent in 1989 to 14 percent in 1995, and grew slowly to 16 percent in 2012.

### 5.1 *The Effects of Mandates on SBP Participation*

Using within-state variation in the SBP mandate policies, I first examine the direct impact of mandates on SBP participation between 1995 and 2012.[15] During this period, eight states began to implement certain types of mandates. Among them, Vermont, Washington D.C. and Rhode Island implemented full coverage mandates that require all school districts to provide the SBP.[16] I employ the following linear probability model to test the effect of the mandate policies on program participation:

$$P_{ist} = \alpha_1 + \theta_1 Mandate_{st} + X_{ist}\beta_1 + Z_{st}\delta_1 + \gamma_s + \mu_t + \lambda_{st} + \epsilon_{ist} \quad (6)$$

where $P$ is a binary variable indicating SBP participation; it equals to one if mother $i$ living in state $s$ in year $t$ has at least one child receiving free/reduced-price breakfast and zero otherwise. $Mandate_{st}$ is an indicator variable equal to one if there is a mandate in state $s$ in year $t$ and zero otherwise; $X$ are maternal demographic characteristics; $Z$ stands for a vector of state-level controls; $\gamma_s$ and $\mu_t$ are state and year fixed effects; $\lambda_{st}$ is a series of state-specific linear time trends, and $\epsilon_{ist}$ is the error term.

The maternal demographic controls include age, race/ethnicity, citizenship, educational attainment, marital status, family size, an indicator of residence in a metropolitan area, number of children, and an indicator of having children younger than five. The state-level controls are annual unemployment rate, per capita personal income, poverty rate and population. Several other state/time varying policies might also affect maternal employment. Therefore, I control for the

---

[15] The SBP participation data was first available in the CPS Food Security Supplement in 1995.

[16] As shown in Figure 3, there is little variation in the percentage of mothers potentially affected by full coverage mandates.

maximum monthly AFDC/ TANF benefits for a three-person family, an indicator for an AFDC waiver, the maximum annual state EITC for a family with two children, and a measure of the generosity of Medicaid or State Children's Health Insurance Program (SCHIP), which is a simulated eligible measure constructed similarly to the one in Currie and Gruber (1996). Using a 1990 national sample, I calculate for each state and each year the percent of infants and children who would be eligible for Medicaid or State Child Health Insurance Programs (SCHIP). This variable varies only by legislative generosity within each state and over time, which does not capture the demographic characteristics of an actual state population that might affect infant health outcomes.

In addition to these control variables, state and year fixed effects are included because unobservable national time trends or cross-sectional state characteristics could bias the results if these factors are correlated with both SBP participation and the mandate policies. For example, national trends in people's attitudes toward child nutrition and each state's generosity in subsidizing the child care could affect SBP participation rate and also influence the passage of mandate policies. The time trends of some factors, such as food prices, may even differ across states. As a result, state-specific linear time trends are also included to account for differences in time trends across states. All estimates are weighted using the CPS supplement weights, and robust standard errors are clustered at the state level.

### 5.2 *The Effects of the School Breakfast Program on Maternal Labor Supply*

Next, making use of the same source of policy variation and data from 1989 to 2012, this paper estimates the effect of the SBP mandates on maternal labor supply.[17] During this period, 21 states initiated SBP mandates and seven states changed the thresholds in their mandates. Six states (Florida, New York, Rhode Island, South Carolina, D.C., and Vermont) adopted full coverage mandates (the mandate threshold of zero) that require all school districts to provide the SBP. A number of mandate implementations and changes during the study period ensure that there is enough power to detect any impact of the SBP mandates on maternal labor supply. I estimate the following model:

$$Y_{ist} = \alpha_2 + \theta_2 Mandate_{st} + X_{ist}\beta_2 + Z_{st}\delta_2 + \gamma_s + \mu_t + \lambda_{st} + \upsilon_{ist} \quad (7)$$

where $Y$ denotes the labor supply outcomes of mother $i$ living in state $s$ at time $t$, and $Mandate_{st}$ refers to the mandate policy in state $s$ at time $t$, which varies by state and year. The estimate of primary interest is $\theta_2$. The only difference between Equation (6) and Equation (7) is that the outcome is maternal labor supply rather than program participation; the control variables $X$, $Z$, $\gamma_s$ and $\mu_t$ are exactly the same.

Three labor supply outcomes are examined: (1) the probability of being employed, (2) the probability of working full time, and (3) weekly hours of work. Outcomes (1) and (2) are estimated using linear probability models for ease of interpretation of estimated marginal effects. For the outcome of weekly hours of work, I use the Heckman (1993) two-step procedure, which generates a consistent estimator when there is non-random selection into maternal employment.[18]

[17] Note that this is a reduced-form analysis. Complete data on employment (used in this analysis) and SBP participation (used in the previous section) are not available during the same CPS month.

[18] The selection equation estimates the probability of being employed on mothers' age, race, citizenship, marital status, number of children, and an indicator for having a child less than five.

## 6. Data

The data used in this paper come from the Current Population Survey (CPS), which is a nationally representative household survey of the U.S. civilian, non-institutionalized population. It surveys approximately 60,000 households each month. The CPS data not only collects a wide range of measures relating to employment and earnings history, but also gathers extensive demographic information from respondents.

Since CPS basic monthly data does not contain school breakfast participation information, I use data from the CPS Food Security Supplement to demonstrate the effect of mandates on program participation. The Food Security dataset has detailed records about food spending and public program participation for each household. The dependent variable is constructed from the survey question "Did any of your children receive free/reduced-price breakfast at school?" This is not a perfect measure of SBP participation since some program participants pay full price. However, given that more than 70 percent of SBP participants receive free/reduced-price breakfast, this measure should be a strong proxy for overall participation rate changes. The Food Security Supplement survey was first conducted in April 1995, and then once a year thereafter. This model is estimated using data from 1995 to 2012. After restricting mothers to those aged 19-50, the final sample contains 251,490 observations.

The maternal labor supply outcomes are drawn from the Annual Social and Economic Supplement (ASEC) of the CPS (also called the March CPS). It contains individuals' employment information for the preceding year, such as labor force participation, average hours of work per week, and hourly wage rate. In models of the impact of mandates on maternal labor supply, I use interviews from 1990 to 2013, indicating a study period of 1989 to 2012. The sample is limited to mothers between 19 and 50 years old who are not in the armed forces. Mothers who were ill or disabled during the previous year or who report positive earnings but zero hours of work are also excluded (0.5 percent of the sample). Additionally, mothers whose children are all over 18 or under 5 are dropped from the sample. I further clean the sample by excluding mothers whose household income is greater than $500,000 (in 2000 dollars) and those who work more than 65 hours per week on average (one percent of the sample in total). The resulting sample size is 433,144 for the March CPS.

Data on the SBP mandates for each state are taken directly from state statues through the *WestLaw* database. The Annual SBP Scorecard Report published by the Food Research and Action Center since 2003 also provides some information about the SBP mandates. Data on AFDC/TANF benefits is obtained from Green Book (U.S. House of Representatives). The AFDC waiver indicator is coded using the Department of Health and Human Services information on the implementation dates for the waivers. The generosity of Medicaid/SCHIP is a simulated measure of public insurance eligibility. It is constructed in a similar way as Currie and Gruber (1996) using policy data from the National Governor's Association. Data on annual state EITC is collected from Center on Budget and Policy priorities. Unemployment data are taken from Bureau of Labor Statistics. Per capita personal income, poverty rate, and population data are downloaded from Bureau of Economic Analysis. I merge the CPS data with mandate policies and state-level controls by state and year.

Table 1 and 2 describe the summary statistics for variables used in this analysis. In addition to the full sample of all mothers, I also present the descriptive statistics for two sub-samples: less-educated mothers (defined as having a high school diploma or less) and single mothers. In the CPS Food Security Supplement full sample, 10 percent of the mothers are black, 43 percent have a high school diploma or less, and 18 percent are single. Mothers who report at least one child receiving free/reduced-price breakfast account for 11.2 percent of the full sample. In the sample of less-educated mothers, 18.1 percent report that their children eat free/reduced-price breakfast. This number is even higher in the sample of single mothers: 22.2 percent. During the period between 1995 and 2012, 50 percent of the mothers lived in a state with a certain type of mandate; 13 percent lived in a state with a mandate requiring all school districts to provide the SBP. The maternal demographic characteristics of the March CPS sample are very similar to those of the Food Security Supplement sample. In the March CPS sample of all mothers, 11 percent are black, 45 percent have a high school diploma or less, 23 percent are single, 77 percent are employed, and 55 percent work full time. The mean of average hours worked per week is 27.49 hours. Compared with the sample of all mothers, the sample of less-educated mothers has a lower employment rate (70 percent) and fewer hours of work per week (24.75 hours). The labor supply of single mothers is relatively high: 80 percent of single mothers are employed, and the average hours worked per week is 30.19 hours. In the March CPS full sample, there are 44 percent of mothers living in a state with some type of mandate.

## 7. Results

### 7.1 *The Impact of Mandates on SBP Participation*

Table 3 displays the participation estimates for Equation (6). Linear probability models are estimated for ease of interpretation.[19] I start by measuring the impact of the SBP mandates in the sample of all mothers (Column 1), and then restrict data to specific subgroups (Columns 2-6). This

[19] Probit models are also estimated for a specification check, and the results are quite robust.

is because SBP participation rate varies widely within each demographic group, and so some groups are more likely to be affected than others. Appendix Table 2 presents SBP participation rates by marital status, education level, race, income, and central city status. These tabulations show that while the participation is widespread across many demographic groups, it is highest among female headed households and low-income families. For example, among single mothers, 22.18 percent have at least one child receiving free/reduced-price school breakfast (CPS data: 1995-2012), and 34.30 percent participate in the School Breakfast Program (SIPP data: 2004). The participation rates are also uniformly higher among mothers with a high school diploma or less, black mothers, and mothers living in central city. In order to increase the power to detect any potential effects of the SBP, I limit the sample to a few subgroups that are more likely to be impacted by the program.

Column 1 of Table 3 shows the results for the full sample of all mothers. There is no significant effect of the SBP mandate on program participation. The results of maternal demographic control variables indicate that, all else equal, mothers who are black or single, or who have a high school diploma or less are more likely to have at least one child receiving subsidized school breakfast. This is consistent with the simple summary statistics of participation rates shown in Appendix Table 2. The likelihood of SBP participation also increases with the number of children. This might be explained by the stigma and fixed cost associated with SBP participation. Except for state population, the state-level controls do not have significant effects on SBP participation. All models include state and year fixed effects and state-specific linear time trends.

Column 2 reports the results for less-educated mothers. In this sample, the implementation of mandates increases the likelihood of having at least one child receiving free/reduced-price breakfast by 1.6 percentage points. Since the SBP participation rate is 18.07 percent among less-educated mothers during the study period, the SBP mandate increases the participation rate by approximately 8.85 percent (1.6/18.07). Among single mothers, the SBP mandate results in a 1.3 percentage point increase in the probability of receiving free/reduced-price breakfast (shown in Column 3). This positive impact of the mandate is even larger in a more targeted subgroup: less-educated single mothers. As displayed in Column 4, the implementation of the SBP mandate increases the probability of receiving a free/reduced-price breakfast by 1.8 percentage points.

It is important to note that this positive association between the mandate and SBP participation does not exist among mothers with a bachelor's degree or above, which is shown in the last column of Table 3. One possible explanation is that most highly educated mothers live in low-poverty districts where state mandates are not effective. Overall, as expected, the state mandate is effective in encouraging children in certain types of families to participate in the SBP.

### 7.2 *The Impact of Mandates on Maternal Labor Supply*

I have shown that the SBP mandate leads to a higher SBP participation rate. This gives me confidence that in the reduced-form labor supply models that follow, SBP mandates are providing enough variation in availability of the SBP. Table 4 presents the results of estimates of Equation (7) for the full sample and select sub-samples; only the coefficients of "Mandate" variable are reported. Here, mandates can be specified in two ways: (1) a dummy variable taking the value of one if the mother lives in state *s* that had a mandate in year *t* and zero otherwise; and (2) two dummy variables, a "full coverage mandate" dummy equal to one if the mother lives in a state with a mandate that requires all school districts to participate in the SBP and zero otherwise, and a "partial coverage mandate" dummy equal to one if the mother lives in a state with a mandate that only asks school districts with a high percentage of students eligible for subsidized breakfast to provide the SBP.[20]

Each cell of Table 4 presents results for one labor supply outcome. Effects on the probability of being employed are displayed in the first cell of Table 4. In the full sample of all mothers, the evidence does not show that the SBP mandate significantly affects maternal employment (shown in Column 1). Among less-educated mothers, there are no significant effects when the mandate variable is specified as "Any Mandate." When the mandate variable enters into the equation as two dummies, the results show that a full coverage mandate is associated with a 2.6 percentage point increase in the probability of being employed, but a weaker partial coverage mandate does not have significant impact on maternal employment (shown in Column 2). Similar to the result for less-educated mothers, Columns 3 and 4 show that a full coverage mandate is associated with a 5.8 and a 5.7 percentage point increase in the probability of being employed for single mothers and less-educated single mothers, respectively. Table 5 displays the results for all other control variables in maternal employment models when the mandate variable is specified as full and partial coverage dummies. Appendix Table 3 presents the full regression results when the mandate variable is specified as "Any Mandate."

I further investigate the impact of mandates on the probability of working full time and display the results in the second panel of Table 4.[21] Among all mothers, the evidence shows that the mandates do not significantly affect mothers' probability of working full time. But among less-educated mothers, single mothers with any level of education, and less-educated single mothers, a

[20] The mandate variable is not specified as a full or partial coverage dummy in the participation equation estimated above because there was not enough variation in full coverage mandates between 1995 and 2012.

[21] In the following models, the full regression results are not shown due to space limitation. Results are available upon request.

significant positive association exists between a full coverage mandate and the likelihood of working full time. Results from Columns 2-4 indicate that a full coverage mandate increases the probability of working full time by 3.4 to 5.9 percentage points for different demographic groups. However, the partial coverage mandates do not have any significant effect in any sample of mothers.

The state mandate may not only affect maternal labor supply at the extensive margin, but also at the intensive margin. Hence, the effects of mandates on mothers' weekly hours of work are also examined and the results are reported in the third panel of Table 4. Again, the evidence shows that the mandates do not affect weekly hours of work in the sample of all mothers. However, in the sub-samples, the findings show that a full coverage mandate is associated with a significant increase in average hours of work per week. A full coverage mandate increases average hours of work per week by approximately 1.25 hours for less-educated mothers, 1.98 hours for single mothers, and 2.31 hours for less-educated single mothers. There is no evidence that partial coverage mandates are significantly associated with hours of work per week.

Overall, the results presented in Table 4 suggest that a full coverage mandate increases the probability of being employed and working full time, and weekly hours of work among less-educated mothers and single mothers.

### 7.3 *Specification Checks*

I perform several specification checks to test the robustness of my results. First, given that linear probability models can be biased and inconsistent if the estimated probabilities are not bounded on the unit level, I estimate probit models for all samples and find that the results are robust to the choice of specification.[22] Second, to verify that these estimates of the relationship between maternal labor supply and the SBP mandates are not spurious, I run several placebo regressions or falsification tests. Given that a large fraction of SBP participants are from low-income or less-educated families, the labor supply effect of the SBP should mainly be focused on mothers from these families. There should not be any significant effect for mothers with higher income or education levels; or at least these effects should be small. If there exists a large and significant effect on mothers whose children are less likely to participate in the program, then the positive effect that I find earlier for the targeted group is unlikely to be a causal one. I estimate Equation (7) using three sub-samples of mothers who are unlikely to be affected by the SBP mandates. These sub-samples are: (1) mothers whose incomes are greater than $100,000 (in 2000 dollars), (2) mothers with a bachelor's degree or above, and (3) married mothers whose household incomes are greater than $50,000. As shown in Table 6, the coefficients of the mandate variables are statistically insignificant across all three sub-samples. There is no evidence that the SBP mandates affect labor supply among mothers whose children are much less likely to participate in the SBP. Therefore, the baseline estimate does not appear to be caused by a spurious correlation between mandates and state labor market trends.

Third, I look at the effect of mandates on mothers' labor force participation. Given that the group of mothers who are in the labor force may include those who want to work but cannot find a job, the labor supply effect of a full coverage mandate should also be positive and might be larger than the effect on employment. As shown in the first panel of Table 7, a full coverage mandate increases the labor force participation by 3.6 and 6.2 percentage points among less-educated

[22] Due to space limitation, the results are not presented in this paper, but are available upon request.

mothers and single mothers, respectively. The magnitude of the effect is slightly larger than the effect on maternal employment.

Fourth, the positive labor supply effects I find suggest that the main mechanism by which the SBP affects labor supply is that mothers redistribute time from household production to paid work. To confirm that this is actually the main mechanism for the labor supply increase, I look at the effects of mandates on other employment outcomes, taking advantage of the fact that the CPS asks respondents why they were not working in the past year. The responses include: (1) "not working because of taking care of home/family," (2) "not working because of going to school," (3) "not working because could not find work for a long time," and (4) "not working because of other reasons not specified." I create three dichotomous outcome variables: the first two are based on responses (1) and (2) respectively, and the third is a combination of responses (3) and (4). As shown in Table 7, among less-educated mothers and single mothers, a full coverage mandate is associated with a reduction in reporting "not working because of taking care of home/family," but there is no significant effect on the other outcomes. The results indicate that availability of the SBP mainly gives mothers more time to participate in the labor force instead of spending time at home engaging in cooking for children.

## 8. Conclusion

By using state/year-level variation in the passage of the SBP mandates, this paper estimates the impact of the SBP availability on maternal labor supply. I first show that the implementation of a mandate significantly increases the probability of children receiving free/reduced-price breakfast in families headed by mothers with a high school diploma or less, or single mothers with any level of education. Then I estimate reduced-form labor supply equations with measures of mandate strength, using March CPS data from 1989 to 2012. In the full sample of all mothers, there is no evidence that mandates are associated with a significant change in maternal labor supply. However, in sub-samples of less-educated mothers and single mothers, the results indicate that a full coverage mandate that requires all school districts to provide the SBP has a positive effect on maternal labor supply. Among less-educated mothers, a full coverage mandate increases the probability of being employed by 2.6 percentage points, the probability of working full time by 3.4 percentage points, and weekly hours of work by 1.25 hours (or 65 hours per year). Among single mothers with any level of education, a full coverage mandate is associated with a 5.8 percentage point increase in the probability of being employed and a 1.98 hour increase in weekly hours of work. The fact that the significant labor supply effect only appears among single mothers but not married mothers is consistent with the literature on child care subsidies and maternal employment, which suggests that employment elasticities are larger for single than for married mothers. The results also show that a partial coverage mandate does not have a significant labor supply effect.

There are several policy implications of these results. The direct goal of the SBP is to improve child nutrition and health, and policy makers may not take maternal employment into consideration as a possible spillover effect. However, this research illustrates that there is a potentially positive effect of the SBP on mothers' employment. Labor supply of women, particularly single mothers, has been a focus of income-support policy reforms in the United States for many years. One goal of welfare reform in the 1990s was to help low-income women, especially welfare recipients, move out of poverty and increase their labor force attachment. My findings show the possibility that certain type of nutrition programs like the SBP can have positive employment effects. More broadly, this study provides insights for policy makers on designing means-tested transfer programs which aim to providing benefits to recipients but at the same time help them attain self-sufficiency.

## LIST of REFERENCES


Anderson, Patricia M., and Philip B. Levine. 2000. "Child Care and Mothers Employment Decisions." In *Finding Jobs: Work and Welfare Reform*, ed. David E. Card and Rebecca M. Blank. New York: Russell Sage Foundation.

Apps, Patricia F., and Ray Rees. 1996. "Labour Supply, Household Production and Intra-family Welfare Distribution." *Journal of Public Economics*, 60(2): 199-219.

Averett, Susan L., H. Elizabeth Peters, and Donald M. Waldman. 1997. "Tax Credits, Labor Supply, and Child Care." *Review of Economics and Statistics*, 29(1): 125-135.

Axelsen, Dan, Dan Friesner, Robert Rosenman, and Hal Snarr. 2007. "Welfare Recipient Work Choice and In-kind Benefits in Washington State." *Applied Economics*, 39(8): 1021-1036.

Baker, Michael, Jonathan Gruber, and Kevin Milligan. 2008. "Universal Childcare, Maternal Labor Supply, and Family Well-being." *Journal of Political Economy*, 116(4): 709-745.

Bartlett, Susan, Frederic Glantz, and Christopher Logan. 2008. "School Lunch and Breakfast Cost Study-II, Final Report." U.S. Department of Agriculture, Food and Nutrition Service, Office of Research, Nutrition and Analysis, Alexandria, VA.

Baum, Charles L. 2002. "A Dynamic Analysis of the Effect of Child Care Costs on the Work Decisions of Low-income Mothers with Infants." *Demography*, 39(1): 139-164.

Becker, Gary S. 1965. "A Theory of the Allocation of Time." *The Economic Journal*, 75(299): 493-517.

Berger, Mark C., and Dan A. Black. 1992. "Child Care Subsidies, Quality of Care, and the Labor Supply of Low-income, Single mothers." *The Review of Economics and Statistics*, 74(4): 635-642.

Bhattacharya, Jayanta, Janet Currie, and Steven J. Haider. 2006. "Breakfast of Champions? The School Breakfast Program and the Nutrition of Children and Families." *Journal of Human Resources*, 41(3): 445-466.

Bianchi, Suzanne M. 2011. "Family Change and Time Allocation in American Families." *The ANNALS of the American Academy of Political and Social Science*, 638(1): 21-44.

Blau, David. 2003. "Child Care Subsidy Programs." *In Means-Tested Transfer Programs in the United States*, ed. Robert A. Moffitt, 443-516. University of Chicago Press.

Blau, David, and Erdal Tekin. 2001. "The Determinants and Consequences of Child Care Subsidy Receipt by Low-income Families." Northwestern University/University of Chicago Joint Center for Poverty Research 213.

Blau, David, and Erdal Tekin. 2007. "The Determinants and Consequences of Child Care Subsidies for Single Mothers in the USA." *Journal of Population Economics*, 20(4): 719-741.

Cherchye, Laurens, Bram De Rock, and Frederic Vermeulen. 2012. "Married with Children: A Collective Labor Supply Model with Detailed Time Use and Intrahousehold Expenditure Information." *The American Economic Review*, 102(7): 3377-3405.

Connelly, Rachel, and Jean Kimmel. 2003. "Marital Status and Full-time/Part-time Work Status in Child Care Choices." *Applied Economics*, 35(7): 761-777.

Currie, Janet. 2003. "U.S. Food and Nutrition Programs." In *Means-tested Transfer Programs in the United States*, ed. Robert A. Moffitt, 199-290. Chicago: University of Chicago Press.

Currie, Janet, and Jonathan Gruber. 1996. "Saving Babies: The efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women." *Journal of Political Economy*, 104(6): 1263-1296.

Danziger, Sheldon, Robert Haveman, and Robert Plotnick. 1981. "How Income Transfer Programs Affect Work, Savings, and the Income Distribution: A Critical Review." *Journal of Economic Literature*, 19(3): 975-1028.

Elliott, Marta, and John Packham. 1998. "When Do Single Mothers Work? An Analysis of 1990 Census Data." *Journal of Sociology and Social Welfare*, 25(1): 39-60.

Fox, Mary Kay, William Hamilton, and Biing-Hwan Lin. 2004. "Effects of Food Assistance and Nutrition Programs on Nutrition and Health Food: Volume 3, Literature Review." U.S. Department of Agriculture, Economic Research Service Research Report 19-3, Washington, DC.

Fraker, Thomas, and Robert Moffitt. 1988. "The Effect of Food Stamps on Labor Supply: A Bivariate Selection Model." *Journal of Public Economics*, 35(1): 25-56.

Frisvold, David. 2012. "Nutrition and Cognitive Achievement: An Evaluation of the School Breakfast Program." Working Paper.

Gelbach, Jonah B. 2002. "Public Schooling for Young Children and Labor Supply." *American Economic Review*, 92(1): 307-322.

Gronau, Reuben. 1977. "Leisure, Home Production and Work-The Theory of the Allocation of Time Revisited." *The Journal of Political Economy*, 85(6): 1099-1123.

Hagstrom, Paul A. 1996. "The Food Stamp Participation and Labor Supply of Married Couples: An Empirical Analysis of Joint Decisions." *The Journal of Human Resources*, 31(2): 383-403.

Hamrick, Karen S., Margaret Andrews, Joanne Guthrie, David Hopkins, and Ket McClelland. 2011. "How Much Time Do Americans Spend on Food?" U.S. Department of Agriculture, Economic Research Service EBI-86.

Hausman, Jerry A. 1980. "The Effect of Wages, Taxes, and Fixed Costs of Women's Labor Force Participation." *The Journal of Public Economics*, 14(2): 161-194.

Heckman, James. 1993. "What Has Been Learned about Labor Supply in the Past Twenty Years?" *American Economic Review*, 83(2): 116-121.

Herbst, Chris M. 2010. "The Labor Supply Effects of Child Care Costs and Wages in the Presence of Subsidies and the Earned Income Tax Credit." *Review of Economics of the Household*, 8(2): 199-230.

Hewins, Jessie, and Madeleine Levin. 2013. "School Breakfast Scorecard 2012." Food Research and Action Center, Washington, DC.

Hoynes, Hilary Williamson. 1996. "Welfare Transfers in Two-parent Families: Labor Supply and Welfare Participation Under AFDC-UP." *Econometrica*, 64(2): 295-332.

Hoynes, Hilary Williamson. 1997. "Work, Welfare, and Family Structure: What Have We Learned." In *Fiscal Policy: Lessons from Economic Research*, ed. Alan J. Auerbach. The MIT Press.

Hoynes, Hilary Williamson, and Diane Whitmore Schanzenbach. 2009. "Consumption Responses to In-kind Transfers: Evidence from the Introduction of the Food Stamp Program." *American Economic Journal: Applied Economics*, 1(4): 109-139.

Hoynes, Hilary Williamson, and Diane Whitmore Schanzenbach. 2012. "Work Incentives and the Food Stamp Program." *Journal of Public Economics*, 96(1): 151-162.

Huffman, Sonya Kostova, and Helen H. Jensen. 2005. "Linkages among Welfare, Food Assistance Programmes and Labor Supply: Evidence from the Survey of Programme Dynamics." *Applied Economics*, 37(10): 1099-1113.

Keane, Michael, and Robert Moffitt. 1998. "A Structural Model of Multiple Welfare Program Participation and Labor Supply." *International Economic Review*, 39(3): 553-589.

Kimmel, Jean. 1995. "The Effectiveness of Child-care Subsidies in Encouraging the Welfare-to-work Transition of Low-income Single Mothers." *American Economic Review*, 85(2): 271-275.

Lefebvre, Pierre, and Philip Merrigan. 2008. "Child-care Policy and the Labor Supply of Mothers with Young Children: A Natural Experiment from Canada." *Journal of Labor Economics*, 26(3): 519-548.

Livermore, Michelle M., and Rebecca S. Powers. 2006. "Employment of Unwed Mothers: The Role of Government and Social Support." *Journal of Family and Economic Issues*, 27(3): 479-494.

Meyer, Bruce D., and Dan T. Rosenbaum. 2001. "Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers." *The Quarterly Journal of Economics*, 116(3): 1063-1104.

Moffitt, Robert. 1983. "An Economic Model of Welfare Stigma." *American Economic Review*, 73(5): 1023-1035.

Moffitt, Robert. 1992. "Incentive Effects of the US Welfare System: A Review." *Journal of Economic Literature*, 30(1): 1-61.

Moffitt, Robert. 1999. "The Effect of Pre-PRWORA Waivers on AFDC Caseloads and Female Earnings, Income, and Labor Force Behavior." In *Economic Conditions and Welfare Reform*, ed. Sheldon H. Danziger, 91-118. Kalamazoo.

Moffitt, Robert. 2002. "Welfare Programs and Labor Supply." In *Handbook of Public Economics. Vol. 4. 1 edition*, ed. Alan J. Auerbach and Martin Feldstein, 2393-2430. Elsevier.

Ollinger, Michael, Katherine Ralston, and Joanne Guthrie. 2011. "School Breakfast and Lunch Costs: Are There Economies of Scale?" Agricultural & Applied Economics Association's 2011 Joint Annual Meeting.

Ribar, David C. 1995. "A Structural Model of Child Care and the Labor Supply of Married Women." *Journal of Labor Economics*, 13(3): 558-597.

Schoeni, Robert F., and Rebecca M. Blank. 2000. "What Has Welfare Reform Accomplished? Impacts on Welfare Participation, Employment, Income, Poverty, and Family Structure." National Bureau of Economic Research, Working Paper No.7276.

Tekin, Erdal. 2005. "Child Care Subsidy Receipt, Employment, and Child Care Choices of Single Mothers." *Economics Letters*, 89(1): 1-6.

Tekin, Erdal. 2007. "Single Mothers Working at Night: Standard Work, Child Care Subsidies, and Implications for Welfare Reform." *Economic Inquiry*, 45(2): 233-250.

Wales, Terence J., and Alan D. Woodland. 1977. "Estimation of the Allocation of Time for Work, Leisure, and Housework." *Econometrica*: *Journal of the Econometric Society*, 45(1): 115-132.

**Figure 1. Consumption-Leisure Response to the In-Kind Transfer of the SBP**

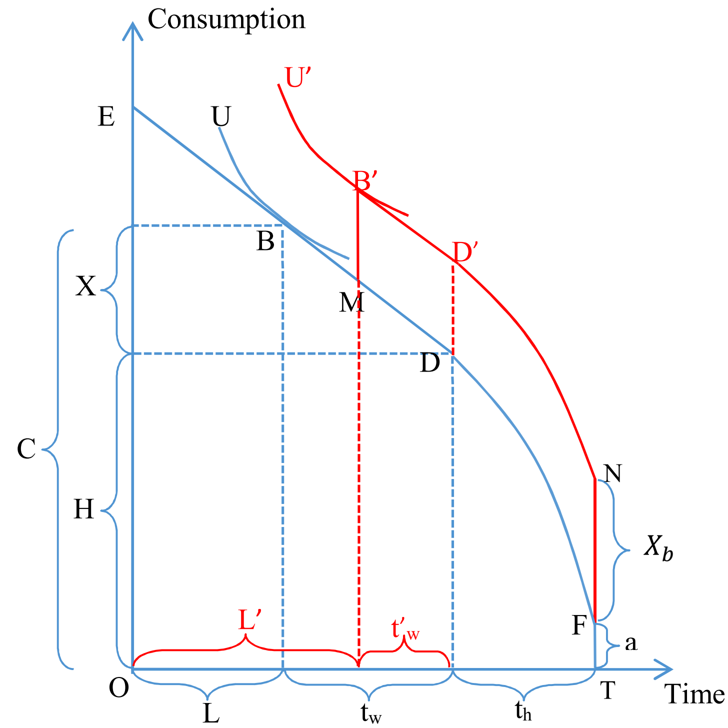


**Figure 2. Consumption-Leisure Response to the Availability of the SBP**

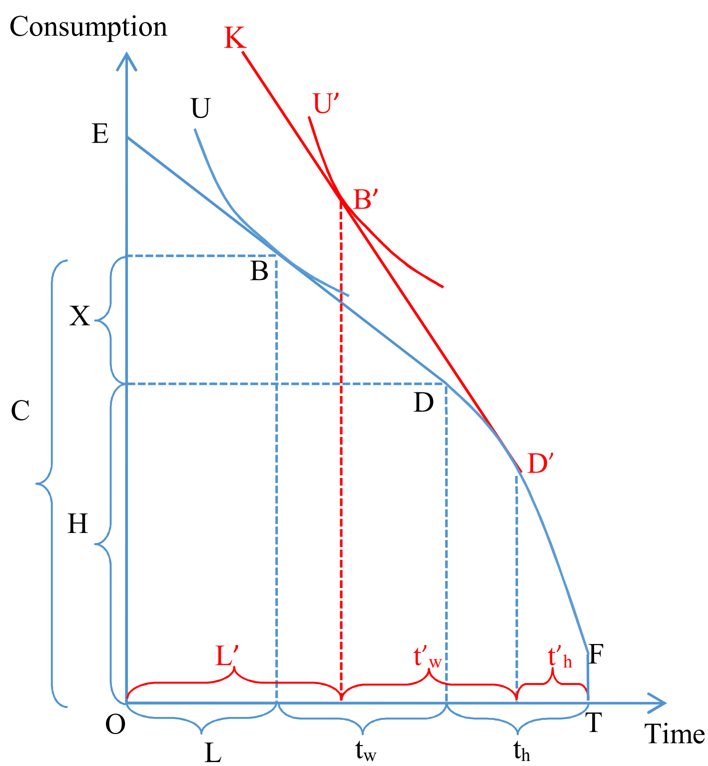
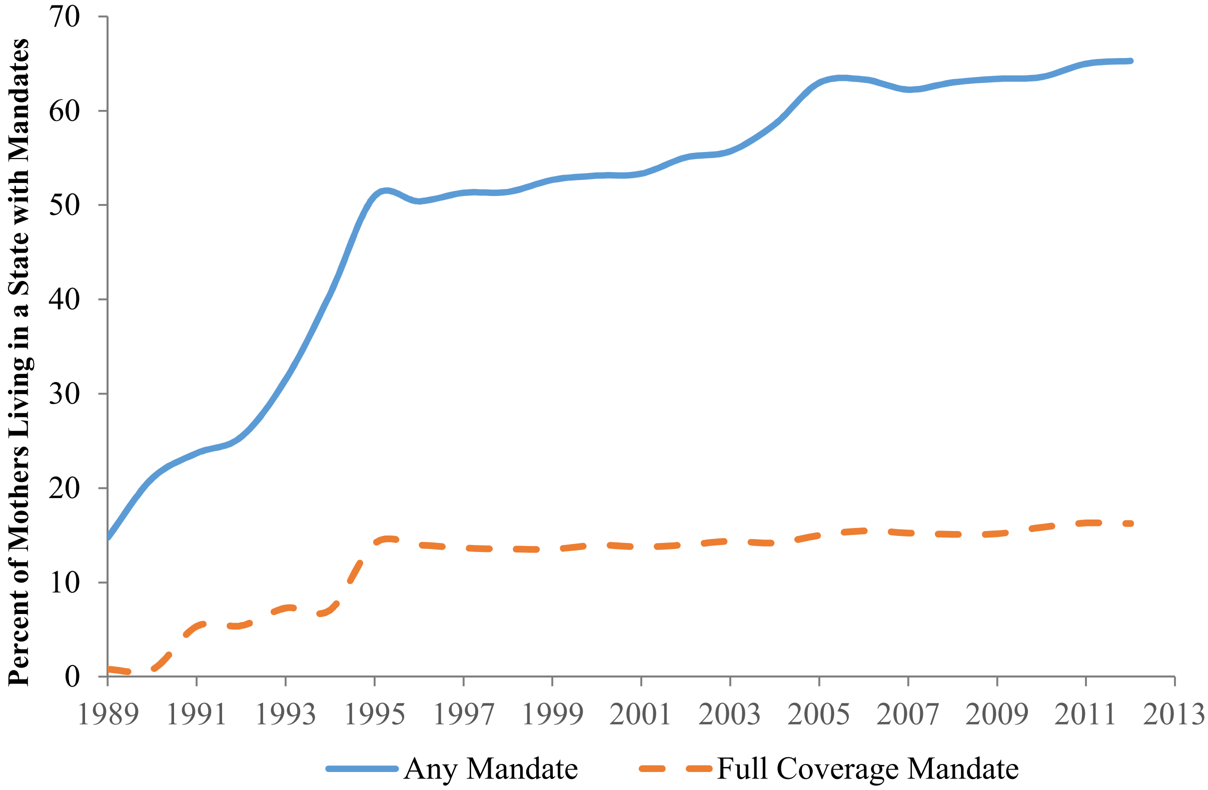


**Figure 3. Percent of Mothers Living in a State with the SBP Mandate**




Source: March CPS data during the period 1990-2013

Table 1. Descriptive Statistics for Demographic Variables (Variable Means)

| Data | | 1995-2012 CPS Food Security Supplement | | | 1990-2013 March CPS | | |
|---|---|---|---|---|---|---|---|
| **Sample** | | **All** | **Less-Educated** | **Single** | **All** | **Less-Educated** | **Single** |
| **Mother Demographic** | | | | | | | |
| Age | | 37.07 | 36.77 | 36.56 | 37.76 | 36.52 | 36.31 |
| | | (7.36) | (6.85) | (7.15) | (6.58) | (6.74) | (7.08) |
| Race | White | 0.84 | 0.82 | 0.70 | 0.83 | 0.82 | 0.69 |
| | | (0.37) | (0.39) | (0.46) | (0.37) | (0.38) | (0.46) |
| | Black | 0.10 | 0.13 | 0.26 | 0.11 | 0.13 | 0.26 |
| | | (0.30) | (0.33) | (0.44) | (0.31) | (0.33) | (0.44) |
| | Other | 0.06 | 0.06 | 0.05 | 0.06 | 0.06 | 0.05 |
| | | (0.24) | (0.23) | (0.21) | (0.24) | (0.23) | (0.21) |
| Education | <9 Years | 0.03 | 0.11 | 0.04 | 0.05 | 0.10 | 0.05 |
| | | (0.18) | (0.31) | (0.20) | (0.21) | (0.30) | (0.21) |
| | 9-12 Years | 0.08 | 0.18 | 0.13 | 0.11 | 0.24 | 0.16 |
| | | (0.27) | (0.39) | (0.33) | (0.31) | (0.42) | (0.37) |
| | High School | 0.32 | 0.71 | 0.34 | 0.29 | 0.61 | 0.32 |
| | | (0.47) | (0.45) | (0.47) | (0.45) | (0.49) | (0.47) |
| | Some College | 0.30 | | 0.34 | 0.29 | | 0.32 |
| | | (0.46) | | (0.47) | (0.45) | | (0.46) |
| | College + | 0.26 | | 0.15 | 0.24 | | 0.14 |
| | | (0.44) | | (0.35) | (0.43) | | (0.34) |
| Hispanic | | 0.10 | 0.29 | 0.20 | 0.18 | 0.26 | 0.20 |
| | | (0.31) | (0.45) | (0.40) | (0.38) | (0.44) | (0.40) |
| Resides in a Central City | | 0.20 | 0.25 | 0.31 | 0.22 | 0.25 | 0.32 |
| | | (0.39) | (0.43) | (0.46) | (0.41) | (0.43) | (0.47) |
| Resides in MSA, Not in a | | 0.51 | 0.51 | 0.49 | 0.55 | 0.50 | 0.48 |
| Central City | | (0.48) | (0.50) | (0.50) | (0.50) | (0.50) | (0.50) |
| U.S. Citizen | | 0.90 | 0.82 | 0.92 | 0.87 | 0.64 | 0.80 |
| | | (0.28) | (0.39) | (0.27) | (0.32) | (0.48) | (0.40) |
| Single | | 0.18 | 0.29 | | 0.23 | 0.28 | |
| | | (0.38) | (0.45) | | (0.42) | (0.45) | |
| Number of Children | | 1.93 | 2.32 | 2.03 | 2.25 | 2.32 | 2.05 |
| | | (0.95) | (1.10) | (1.03) | (1.03) | (1.10) | (1.05) |
| Have a Child <5 | | | | | 0.26 | 0.27 | 0.21 |
| | | | | | (0.44) | (0.44) | (0.41) |
| Total Family Income/1000 | | 60.03 | 41.78 | 26.46 | 59.60 | 41.73 | 25.25 |
| | | (53.99) | (38.38) | (26.88) | (53.87) | (36.39) | (25.47) |
| N | | 251,490 | 108,140 | 45,269 | 433,144 | 203,398 | 101,456 |

*Notes: Summary statistics for mothers aged 19-50. Standard deviations are in parenthesis. In March CPS sample, mothers who are in the armed forces or cannot work because of sick or disabled are deleted. Mothers who do not have at least one child between 5 and 18 are also excluded from the sample.*
*Source: CPS Food Security Supplement during the period 1995-2012 and March CPS from 1990 to 2013*

**Table 2. Descriptive Statistics for Outcome and Policy Variables and State Controls**

| Data | 1995-2012 CPS Food Security Supplement | | | 1990-2013 March CPS | | |
|---|---|---|---|---|---|---|
| **Sample** | **All** | **Less- Educated** | **Single** | **All** | **Less-Educated** | **Single** |
| **Outcome Variables** | | | | | | |
| Any Child Receives Free/ | 0.11 | 0.18 | 0.22 | | | |
| Reduced-Price Breakfast | (0.31) | (0.35) | (0.34) | | | |
| Employed | | | | 0.77 | 0.70 | 0.80 |
| | | | | (0.42) | (0.46) | (0.40) |
| Work Full Time | | | | 0.55 | 0.50 | 0.65 |
| | | | | (0.50) | (0.50) | (0.48) |
| Average Hours/Week | | | | 27.49 | 24.75 | 30.19 |
| | | | | (17.60) | (18.29) | (16.91) |
| **Mandate Policies** | | | | | | |
| Any Mandate | 0.50 | 0.49 | 0.52 | 0.44 | 0.41 | 0.47 |
| | (0.50) | (0.50) | (0.50) | (0.50) | (0.49) | (0.50) |
| Full Coverage Mandate | 0.13 | 0.14 | 0.16 | 0.12 | 0.11 | 0.14 |
| | (0.34) | (0.35) | (0.37) | (0.32) | (0.31) | (0.34) |
| Partial Coverage Mandate | 0.36 | 0.35 | 0.36 | 0.33 | 0.3 | 0.34 |
| | (0.48) | (0.48) | (0.48) | (0.47) | (0.46) | (0.47) |
| **State Level Controls** | | | | | | |
| Maximum AFDC/TANF for a | 397.94 | 401.85 | 397.20 | 421.68 | 430.49 | 412.61 |
| 3-Person Family | (150.11) | (155.28) | (152.31) | (169.34) | (179.99) | (170.33) |
| Maximum Annual State EITC | 263.17 | 217.71 | 249.32 | 215.21 | 171.29 | 219.55 |
| for a Family with 2 Children | (489.07) | (447.60) | (475.56) | (451.32) | (405.04) | (453.24) |
| Medicaid/SCHIP Generosity | 48.48 | 46.55 | 47.56 | 42.13 | 38.26 | 42.70 |
| | (13.45) | (13.63) | (13.47) | (18.21) | (19.57) | (17.98) |
| AFDC Waiver | 0.97 | 0.96 | 0.97 | 0.79 | 0.71 | 0.80 |
| | (0.17) | (0.19) | (0.18) | (0.41) | (0.45) | (0.40) |
| Unemployment Rate | 5.77 | 5.51 | 5.60 | 5.88 | 5.90 | 6.02 |
| | (2.05) | (1.83) | (1.90) | (1.97) | (1.89) | (1.99) |
| Per Capita Personal | 30.94 | 30.42 | 30.81 | 29.82 | 29.15 | 30.05 |
| Income/1000 | (4.77) | (4.68) | (5.01) | (5.06) | (5.06) | (5.30) |
| Poverty Rate | 12.56 | 12.66 | 12.70 | 12.86 | 13.17 | 13.20 |
| | (3.21) | (3.20) | (3.25) | (3.33) | (3.37) | (3.41) |
| Population/1000,000 | 10.07 | 10.85 | 10.05 | 9.98 | 10.64 | 10.14 |
| | (10.15) | (10.52) | (9.82) | (9.85) | (10.14) | (9.70) |
| N | 251,490 | 108,140 | 45,269 | 433,144 | 203,398 | 101,456 |

*Notes: Standard deviations are in parentheses. Outcome variables are from CPS data. Mandate polices and state level controls are merged with CPS data by year and by state. Mandate policies are from state statues. AFDC/TANF is from Green Book (U.S. House of Representatives). EITC is from Center on Budget and Policy priorities. Medicaid/SCHIP is constructed the same way as Currie and Gruber (1996). AFDC waiver is coded using the Department of Health and Human Services information on the implementation dates for the waivers. Unemployment data is from Bureau of Labor Statistics. Per Capita personal income, poverty, and population data are from Bureau of Economic Analysis.*

## Table 3. Impacts of Mandates on SBP Participation, by Group

| | All | Less-Educated | Single | Less-Educated &Single | College+ |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| **Any Mandate** | **0.0094** | **0.0159**** | **0.0129*** | **0.0178**** | **-0.0094** |
| | **(0.0059)** | **(0.0066)** | **(0.0066)** | **(0.0068)** | **(0.0152)** |
| Age | 0.0004 | 0.0003 | 0.0009** | 0.0008 | -0.0025 |
| | (0.0004) | (0.0004) | (0.0005) | (0.0005) | (0.0022) |
| Age Square | -0.0000* | -0.0000 | -0.0000** | -0.0000* | 0.0000 |
| | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| Black | 0.0152*** | 0.0131*** | 0.0152*** | 0.0126*** | 0.0195*** |
| | (0.0018) | (0.0020) | (0.0020) | (0.0022) | (0.0051) |
| Other Race | 0.0110*** | 0.0083** | 0.0127*** | 0.0110*** | 0.0041 |
| | (0.0035) | (0.0036) | (0.0037) | (0.0037) | (0.0055) |
| Education < 9 Years | 0.0082*** | 0.0081*** | 0.0076*** | 0.0070*** | |
| | (0.0012) | (0.0011) | (0.0012) | (0.0011) | |
| Education 9-12 Years | 0.0059*** | 0.0060*** | 0.0051*** | 0.0052*** | |
| | (0.0007) | (0.0007) | (0.0010) | (0.0010) | |
| Some College | -0.0046*** | | -0.0058*** | | |
| | (0.0008) | | (0.0010) | | |
| College+ | -0.0110*** | | -0.0118*** | | |
| | (0.0019) | | (0.0025) | | |
| Hispanic | 0.0084*** | 0.0070*** | 0.0078*** | 0.0050*** | 0.0088 |
| | (0.0012) | (0.0013) | (0.0016) | (0.0016) | (0.0054) |
| Resides in a Central City | -0.0025 | -0.0013 | -0.0022 | -0.0001 | 0.0015 |
| | (0.0020) | (0.0022) | (0.0021) | (0.0024) | (0.0057) |
| Single | 0.0034*** | 0.0042*** | | | 0.0034 |
| | (0.0007) | (0.0008) | | | (0.0041) |
| U.S. Citizen | -0.0049** | -0.0058** | -0.0066** | -0.0082** | -0.0051 |
| | (0.0024) | (0.0027) | (0.0028) | (0.0033) | (0.0042) |
| Number of Children | 0.0052*** | 0.0057*** | 0.0077*** | 0.0078*** | 0.0021*** |
| | (0.0015) | (0.0016) | (0.0018) | (0.0032) | (0.0012) |
| AFDC/TANF Benefits | 0.0000 | 0.0000 | 0.0000 | -0.0000 | 0.0001 |
| | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0001) |
| EITC Benefits | 0.0000 | -0.0000 | 0.0000 | -0.0000 | 0.0000 |
| | (0.0000) | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| Medicare/SCHIP | -0.0001 | -0.0001 | -0.0000 | -0.0000 | -0.0000 |
| Generosity | (0.0001) | (0.0001) | (0.0001) | (0.0001) | (0.0003) |
| AFDC Waiver | -0.0023 | -0.0025 | -0.0026 | -0.0005 | 0.0053 |
| | (0.0020) | (0.0023) | (0.0025) | (0.0028) | (0.0105) |
| Unemployment Rate | -0.0004 | -0.0012 | -0.0007 | -0.0022 | 0.0007 |
| | (0.0011) | (0.0013) | (0.0011) | (0.0015) | (0.0036) |
| Per Capita Personal | 0.0005 | 0.0001 | 0.0005 | -0.0004 | -0.0040 |
| Income/1,000 | (0.0007) | (0.0008) | (0.0009) | (0.0011) | (0.0032) |
| Poverty Rate | -0.0001 | -0.0003 | -0.0003 | -0.0003 | 0.0009 |
| | (0.0003) | (0.0004) | (0.0004) | (0.0004) | (0.0014) |
| Population/1,000,000 | -0.0015* | -0.0018* | -0.0003 | 0.0010 | 0.0019 |
| | (0.0008) | (0.0010) | (0.0009) | (0.0012) | (0.0042) |
| Constant | 0.0522** | 0.0818** | 0.0389 | 0.0708* | 0.1273 |
| | (0.0252) | (0.0314) | (0.0322) | (0.0412) | (0.1132) |
| N | 251,490 | 108,140 | 45,269 | 31,735 | 65,387 |

*Notes: 1. Robust standard errors are in parenthesis. ***, **, and * indicate the estimates are statistically significant at the 1%, 5%, and 10% levels, respectively. 2. All the models control for state and year fixed effects and state-specific linear time trends. 3. Estimates are weighted using CPS weight, and standard errors are clustered at the state level. Source: CPS Food Security Supplement 1995-2012*

## Table 4. Impacts of Mandates on Maternal Labor Supply, by Group

| | | All (1) | Less-Educated (2) | Single (3) | Less-Educated & Single (4) |
|---|---|---|---|---|---|
| **Outcome: Employed** | | | | | |
| Spec. 1 | Any Mandate | 0.0020 | 0.0041 | 0.0033 | 0.0010 |
| | | (0.0071) | (0.0132) | (0.0092) | (0.0119) |
| Spec. 2 | Full Coverage Mandate | 0.0192 | 0.0263*** | 0.0579** | 0.0573** |
| | | (0.0154) | (0.0104) | (0.0241) | (0.0220) |
| | Partial Coverage Mandate | -0.0032 | -0.0019 | -0.0138 | -0.0198 |
| | | (0.0061) | (0.0155) | (0.0112) | (0.0165) |
| **Outcome: Full Time** | | | | | |
| Spec. 1 | Any Mandate | 0.0002 | 0.0059 | 0.0082 | 0.0119 |
| | | (0.0077) | (0.01312) | (0.0100) | (0.0132) |
| Spec. 2 | Full Coverage Mandate | 0.0201 | 0.0339*** | 0.0527*** | 0.0591*** |
| | | (0.0130) | (0.0161) | (0.0138) | (0.0157) |
| | Partial Coverage Mandate | -0.0062 | -0.0022 | -0.0051 | -0.0042 |
| | | (0.0066) | (0.0140) | (0.0123) | (0.0133) |
| **Outcome: Hours of Work** | | | | | |
| Spec. 1 | Any Mandate | 0.0431 | 0.2267 | 0.0966 | 0.1541 |
| | | (0.2789) | (0.5103) | (0.3288) | (0.4438) |
| Spec. 2 | Full Coverage Mandate | 0.7522 | 1.2481*** | 1.9752** | 2.3138** |
| | | (0.5984) | (0.4411) | (0.8002) | (0.6455) |
| | Partial Coverage Mandate | -0.1625 | -0.0490 | -0.4988 | -0.5466 |
| | | (0.2461) | (0.5662) | (0.4185) | (0.5662) |
| | N | 433,144 | 203,398 | 101,456 | 54,658 |

*Notes: 1. Robust standard errors are in parenthesis. ***, **, and * indicate the estimates are statistically significant at the 1%, 5%, and 10% levels, respectively. 2. Each parameter is from a separate regression of the outcome variable on a specification of the SBP mandate. 3. Demographic controls include mother's age, age squared, race, ethnicity, central city status, education level, citizenship, marital status, number of children, and an indicator for having a child less than 5. 4. State level controls include a measure of the generosity of Medicaid and SCHIP, the maximum AFDC/TANF benefit, an indicator for an AFDC waiver, the state EITC benefit, unemployment rate, per capita personal income, population, and poverty rate. 5. All the models control for state and year fixed effects and state-specific linear time trends. 6. Estimates are weighted using CPS weight, and standard errors are clustered at the state level.*

*Source: March CPS 1990-2013*

**Table 5. Impacts of Mandates on Maternal Employment (Full Results), by Group**

| Outcome: Employed | All (1) | Less-Educated (2) | Single (3) | Less-Educated &Single (4) |
|---|---|---|---|---|
| **Full Coverage Mandate** | **0.0192** | **0.0263*****| **0.0579**** | **0.0573**** |
| | **(0.0154)** | **(0.0104)** | **(0.0241)** | **(0.0220)** |
| **Partial Coverage Mandate** | **-0.0032** | **-0.0019** | **-0.0138** | **-0.0198** |
| | **(0.0061)** | **(0.0155)** | **(0.0112)** | **(0.0165)** |
| Age | 0.0209*** | 0.0247*** | 0.0214*** | 0.0251*** |
| | (0.0016) | (0.0020) | (0.0025) | (0.0035) |
| Age Square | -0.0003*** | -0.0003*** | -0.0003*** | -0.0003*** |
| | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| Black | 0.0339*** | 0.0124 | -0.0219*** | -0.0307*** |
| | (0.0077) | (0.0112) | (0.0066) | (0.0101) |
| Education < 9 Years | -0.1607*** | -0.1539*** | -0.1855*** | -0.1879*** |
| | (0.0114) | (0.0129) | (0.0171) | (0.0171) |
| Education 9-12 Years | -0.1162*** | -0.1272*** | -0.1589*** | -0.1540*** |
| | (0.0042) | (0.0057) | (0.0056) | (0.0060) |
| Some College | 0.0483*** | | 0.0581*** | |
| | (0.0030) | | (0.0047) | |
| College + | 0.0668*** | | 0.1068*** | |
| | (0.0046) | | (0.0052) | |
| Hispanic | 0.0101 | 0.0038 | -0.0084 | -0.0149 |
| | (0.0084) | (0.0105) | (0.0140) | (0.0176) |
| Resides in a Central City | -0.0438*** | -0.0278** | -0.0146 | -0.0199 |
| | (0.0107) | (0.0138) | (0.0110) | (0.0140) |
| Single | 0.0561*** | 0.0674*** | | |
| | (0.0067) | (0.0095) | | |
| U.S. Citizen | 0.0982*** | 0.0770*** | -0.0359*** | -0.0597*** |
| | (0.0079) | (0.0138) | (0.0120) | (0.0148) |
| Number of Children | -0.0423*** | -0.0352*** | -0.0269*** | -0.0291*** |
| | (0.0009) | (0.0014) | (0.0021) | (0.0023) |
| Any Child < 5 | -0.1059*** | -0.1105*** | -0.0709*** | -0.0827*** |
| | (0.0031) | (0.0035) | (0.0056) | (0.0079) |
| AFDC/TANF Benefits | -0.0001*** | -0.0001* | -0.0002*** | -0.0002*** |
| | (0.0000) | (0.0001) | (0.0001) | (0.0001) |
| EITC Benefits | 0.0000** | 0.0000* | 0.0000 | 0.0000 |
| | (0.0000) | (0.0000) | (0.0000) | (0.0000) |
| Medicare/SCHIP | 0.0003** | 0.0002 | 0.0004 | 0.0005 |
| Generosity | (0.0001) | (0.0003) | (0.0004) | (0.0005) |
| AFDC Waiver | 0.0061 | 0.0035 | -0.0030 | -0.0079 |
| | (0.0043) | (0.0076) | (0.0143) | (0.0170) |
| Unemployment Rate | -0.0018 | 0.0003 | -0.0060** | -0.0051 |
| | (0.0017) | (0.0025) | (0.0026) | (0.0039) |
| Per Capita Personal | 0.0011 | 0.0018 | -0.0000 | -0.0012 |
| Income/1,000 | (0.0014) | (0.0015) | (0.0021) | (0.0025) |
| Poverty Rate | -0.0026*** | -0.0067*** | -0.0041*** | -0.0063*** |
| | (0.0007) | (0.0012) | (0.0011) | (0.0017) |
| Population | 0.0014 | 0.0043*** | 0.0051 | 0.0081** |
| | (0.0016) | (0.0013) | (0.0036) | (0.0039) |
| N | 433,144 | 203,398 | 101,456 | 54,658 |

*Notes: 1. Robust standard errors are in parenthesis. ***, **, and * indicate the estimates are statistically significant at the 1%, 5%, and 10% levels, respectively. 2. All the models control for state and year fixed effects and state-specific linear time trends. 3. Due to space limitation, the coefficients of other race and constant are not reported. 4. Estimates are weighted using CPS weight, and standard errors are clustered at the state level.*
*Source: March CPS 1990-2013*

**Table 6. Impacts of Mandates on Maternal Labor Supply, Falsification Test**

| | | Income>100K<br>(1) | College+<br>(2) | Married & Income>50K<br>(3) |
|---|---|---|---|---|
| **Outcome: Employed** | | | | |
| Spec. 1 | Any Mandate | -0.0071 | -0.0009 | 0.0103 |
| | | (0.0097) | (0.0163) | (0.0100) |
| Spec. 2 | Full Coverage Mandate | -0.0051 | 0.0218 | 0.0039 |
| | | (0.0168) | (0.0311) | (0.0168) |
| | Partial Coverage Mandate | -0.0074 | -0.0082 | 0.0107 |
| | | (0.0103) | (0.014) | (0.011) |
| **Outcome: Full Time** | | | | |
| Spec. 1 | Any Mandate | -0.0092 | -0.0038 | -0.0059 |
| | | (0.0087) | (0.0162) | (0.0077) |
| Spec. 2 | Full Coverage Mandate | 0.0208 | -0.0044 | -0.0065 |
| | | (0.0131) | (0.0332) | (0.0128) |
| | Partial Coverage Mandate | -0.0162 | -0.0041 | -0.0047 |
| | | (0.0095) | (0.0162) | (0.0081) |
| **Outcome: Hours of Work** | | | | |
| Spec. 1 | Any Mandate | -0.3378 | -0.1456 | 0.1381 |
| | | (0.3800) | (0.5782) | (0.3427) |
| Spec. 2 | Full Coverage Mandate | 0.4321 | 0.0805 | -0.1014 |
| | | (0.6512) | (1.4146) | (0.6212) |
| | Partial Coverage Mandate | -0.5094 | -0.2200 | 0.2017 |
| | | (0.3632) | (0.5788) | (0.3804) |
| | N | 59,699 | 106,117 | 196,976 |

*Notes: 1. Robust standard errors are in parenthesis. ***, **, and * indicate the estimates are statistically significant at the 1%, 5%, and 10% levels, respectively. 2. Each parameter is from a separate regression of the outcome variable on a specification of the SBP mandate. 3. Demographic controls include mother's age, age squared, race, ethnicity, central city status, education level, citizenship, marital status, number of children, and an indicator for having a child less than 5. 4. State level controls include a measure of the generosity of Medicaid and SCHIP, the maximum AFDC/TANF benefit, an indicator for an AFDC waiver, state EITC benefit, unemployment rate, per capita personal income, population, and poverty rate. 5. All the models control for state and year fixed effects and state-specific linear time trends. 6. Estimates are weighted using CPS weight, and standard errors are clustered at the state level.*
*Source: March CPS 1990-2013*

## Table 7. Robustness/Sensitivity Test

| | | All (1) | Less-Educated (2) | Single (3) | Less-Educated & Single (4) |
|---|---|---|---|---|---|
| **Outcome: Labor Force Participation** | | | | | |
| Spec. 1 | Any Mandate | 0.0052 | 0.0073 | 0.0031 | 0.0007 |
| | | (0.0056) | (0.0122) | (0.0078) | (0.0123) |
| Spec. 2 | Full Coverage Mandate | 0.0213 | 0.0361*** | 0.0623** | 0.0672** |
| | | (0.0151) | (0.0101) | (0.0300) | (0.0267) |
| | Partial Coverage Mandate | -0.0012 | -0.0011 | -0.0162 | -0.0202 |
| | | (0.0056) | (0.0155) | (0.0154) | (0.0201) |
| **Outcome: Not Working Because of Taking Care of Home/Family** | | | | | |
| Spec. 1 | Any Mandate | -0.0036 | -0.0078 | -0.0035 | -0.0052 |
| | | (0.0070) | (0.0131) | (0.0067) | (0.0104) |
| Spec. 2 | Full Coverage Mandate | -0.0241 | -0.0357*** | -0.0613*** | -0.07454*** |
| | | (0.0152) | (0.0110) | (0.0242) | (0.0263) |
| | Partial Coverage Mandate | 0.0020 | -0.0011 | 0.0140 | 0.0170 |
| | | (0.0061) | (0.0145) | (0.0142) | (0.0180) |
| **Outcome: Not Working Because of Going to School** | | | | | |
| Spec. 1 | Any Mandate | -0.0012 | -0.0009 | -0.0017 | -0.0018 |
| | | (0.0010) | (0.0014) | (0.0032) | (0.0038) |
| Spec. 2 | Full Coverage Mandate | 0.0000 | -0.0011 | -0.0028 | 0.0014 |
| | | (0.0012) | (0.0026) | (0.0052) | (0.0051) |
| | Partial Coverage Mandate | -0.0010 | -0.0011 | -0.0021 | -0.0022 |
| | | (0.0013) | (0.0007) | (0.003) | (0.0045) |
| **Outcome: Not Working Because of Reasons Other than Taking Care of Home/Family** | | | | | |
| Spec. 1 | Any Mandate | 0.0008 | 0.0032 | 0.0010 | 0.0062 |
| | | (0.0011) | (0.0024) | (0.0055) | (0.0083) |
| Spec. 2 | Full Coverage Mandate | 0.0031 | 0.0104 | 0.0040 | 0.0141 |
| | | (0.0020) | (0.0056) | (0.0041) | (0.0082) |
| | Partial Coverage Mandate | 0.0000 | 0.0019 | 0.0000 | 0.0022 |
| | | (0.0010) | (0.0021) | (0.0059) | (0.0100) |
| | N | 433,144 | 203,398 | 101,456 | 54,658 |

*Notes: 1. Robust standard errors are in parenthesis. ***, **, and * indicate the estimates are statistically significant at the 1%, 5%, and 10% levels, respectively. 2. Each parameter is from a separate regression of the outcome variable on a specification of the SBP mandate. 3. Demographic controls include mother's age, age squared, race, ethnicity, central city status, education level, citizenship, marital status, number of children, and an indicator for having a child less than 5. 4. State level controls include a measure of the generosity of Medicaid and SCHIP, the maximum AFDC/TANF benefit, an indicator for an AFDC waiver, state EITC benefit, unemployment rate, per capita personal income, population, and poverty rate. 5. All the models control for state and year fixed effects and state-specific linear time trends. 6. Estimates are weighted using CPS weight, and standard errors are clustered at the state level.*

*Source: March CPS 1990-2013*

**Appendix Table 1. The Mandated Thresholds for the School Breakfast Program, 1989-2012**

| State | 2012 | 2011 | 2010 | 2009 | 2008 | 2007 | 2006 | 2005 | 2004 | 2003 | 2002 | 2001 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Alabama | . | . | . | . | . | . | . | . | . | . | . | . |
| Alaska | . | . | . | . | . | . | . | . | . | . | . | . |
| Arizona | . | . | . | . | . | . | . | . | . | . | . | . |
| Arkansas | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |
| California | . | . | . | . | . | . | . | . | . | . | . | . |
| Colorado | . | . | . | . | . | . | . | . | . | . | . | . |
| Connecticut | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 |
| Delaware | . | . | . | . | . | . | . | . | . | . | . | . |
| D.C. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ***0*** | . | . | . | . |
| Florida | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Georgia | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 |
| Hawaii | . | . | . | . | . | . | . | . | . | . | . | . |
| Idaho | . | . | . | . | . | . | . | . | . | . | . | . |
| Illinois | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | ***0.4*** | . | . | . | . |
| Indiana | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | ***0.15*** | ***0.25*** | 0.25 | 0.25 | 0.25 | ***0.25*** | . |
| Iowa | . | . | . | . | . | . | . | . | . | . | . | . |
| Kansas | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 |
| Kentucky | . | . | . | . | . | . | . | . | . | . | . | . |
| Louisiana | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 |
| Maine | . | . | . | . | . | . | . | . | . | . | . | . |
| Maryland | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | 0.15 | ***0.15*** | . |
| Massachusetts | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 |
| Michigan | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |
| Minnesota | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 |
| Mississippi | . | . | . | . | . | . | . | . | . | . | . | . |
| Missouri | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 |
| Montana | . | . | . | . | . | . | . | . | . | . | . | . |
| Nebraska | . | . | . | . | . | . | . | . | . | . | . | . |
| Nevada | . | . | . | . | . | . | . | . | . | . | . | . |
| New Hampshire | . | . | . | . | . | . | . | . | . | . | . | . |
| New Jersey | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | . | . | . |
| New Mexico | 0.85 | 0.85 | 0.85 | . | . | . | . | . | . | . | . | . |
| New York | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| North Carolina | . | . | . | . | . | . | . | . | . | . | . | . |
| North Dakota | . | . | . | . | . | . | . | . | . | . | . | . |
| Ohio | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 |
| Oklahoma | . | . | . | . | . | . | . | . | . | . | . | . |
| Oregon | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 |
| Pennsylvania | . | . | . | . | . | . | . | . | . | . | . | . |
| Rhode Island | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| South Carolina | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| South Dakota | . | . | . | . | . | . | . | . | . | . | . | . |
| Tennessee | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 |
| Texas | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| Utah | . | . | . | . | . | . | . | . | . | . | . | . |
| Vermont | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | . | . |
| Virginia | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 |
| Washington | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 |
| West Virginia | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Wisconsin | . | . | . | . | . | . | . | . | . | . | . | . |
| Wyoming | . | . | . | . | . | . | . | . | . | . | . | . |

Notes: The mandate threshold in the table means that schools must participate in the SBP if the fraction of students eligible for free or reduced price meal is equal to or greater than the threshold in the table. The number "0" indicates the SBP is mandatory in that state. The state without mandate is shown by "."
Source: the mandates are taken directly from state statues.

**Cont. The Mandated Thresholds for the School Breakfast Program**

| | 2000 | 1999 | 1998 | 1997 | 1996 | 1995 | 1994 | 1993 | 1992 | 1991 | 1990 | 1989 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Alabama | . | . | . | . | . | . | . | . | . | . | . | . |
| Alaska | . | . | . | . | . | . | . | . | . | . | . | . |
| Arizona | . | . | . | . | . | . | . | . | . | . | . | . |
| Arkansas | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | ***0.2*** | ***0.35*** | ***0.4*** | . | . |
| California | . | . | . | . | . | . | . | . | . | . | . | . |
| Colorado | . | . | . | . | . | . | . | . | . | . | . | . |
| Connecticut | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.8 | ***0.8*** | . | . | . | . |
| Delaware | . | . | . | . | . | . | . | . | . | . | . | . |
| D.C. | . | . | . | . | . | . | . | . | . | . | . | . |
| Florida | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ***0*** | ***0.4*** | . |
| Georgia | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | ***0.25*** | . | . | . | . | . | . |
| Hawaii | . | . | . | . | . | . | . | . | . | . | . | . |
| Idaho | . | . | . | . | . | . | . | . | . | . | . | . |
| Illinois | . | . | . | . | . | . | . | . | . | . | . | . |
| Indiana | . | . | . | . | . | . | . | . | . | . | . | . |
| Iowa | . | . | . | . | . | . | . | . | . | . | . | . |
| Kansas | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | ***0.35*** | . | . | . | . |
| Kentucky | . | . | . | . | . | . | . | . | . | . | . | . |
| Louisiana | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | ***0.25*** | . | . |
| Maine | . | . | . | . | . | . | . | . | . | . | . | . |
| Maryland | . | . | . | . | . | . | . | . | . | . | . | . |
| Massachusetts | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | ***0.4*** | . | . | . | . |
| Michigan | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2[1] |
| Minnesota | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | ***0.33*** | 0.4 | 0.4 | 0.4 | 0.4 | ***0.4*** |
| Mississippi | . | . | . | . | . | . | . | . | . | . | . | . |
| Missouri | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | ***0.35*** | . | . | . | . |
| Montana | . | . | . | . | . | . | . | . | . | . | . | . |
| Nebraska | . | . | . | . | . | . | . | . | . | . | . | . |
| Nevada | . | . | . | . | . | . | . | . | . | . | . | . |
| New Hampshire | . | . | . | . | . | . | . | . | . | . | . | . |
| New Jersey | . | . | . | . | . | . | . | . | . | . | . | . |
| New Mexico | . | . | . | . | . | . | . | . | . | . | . | . |
| New York | 0 | 0 | 0 | 0 | 0 | ***0*** | ***0.4*** | . | . | . | . | . |
| North Carolina | . | . | . | . | . | . | . | . | . | . | . | . |
| North Dakota | . | . | . | . | . | . | . | . | . | . | . | . |
| Ohio | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33 | 0.33[2] |
| Oklahoma | . | . | . | . | . | . | . | . | . | . | . | . |
| Oregon | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | ***0.25*** | . | . | . |
| Pennsylvania | . | . | . | . | . | . | . | . | . | . | . | . |
| Rhode Island | ***0*** | ***0.2*** | ***0.4*** | . | . | . | . | . | . | . | . | . |
| South Carolina | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ***0*** | ***0.4*** | 0.4 | 0.4 | 0.4[3] |
| South Dakota | . | . | . | . | . | . | . | . | . | . | . | . |
| Tennessee | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25[4] |
| Texas | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | ***0.1*** | . | . | . | . | . | . |
| Utah | . | . | . | . | . | . | . | . | . | . | . | . |
| Vermont | . | . | . | . | . | . | . | . | . | . | . | . |
| Virginia | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | ***0.25*** | . | . | . | . | . |
| Washington | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | ***0.4*** | . |
| West Virginia | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0[5] |
| Wisconsin | . | . | . | . | . | . | . | . | . | . | . | . |
| Wyoming | . | . | . | . | . | . | . | . | . | . | . | . |

Notes: 1. The mandate in MI began in 1980. 2. The mandate in OH began in 1977. 3. The mandate in SC began in 1976. 4. The mandate in TN began in 1986. 5. The mandate in WV began in 1981.
Source: the mandates are taken directly from state statues.

**Appendix Table 2. The SBP Participation Rates by Demographic Group**

| Data | | CPS 1995-2012 | SIPP 2004 Panel |
|---|---|---|---|
| | | **Any Child Eat Free or Reduced-Price Breakfasts** | **Any Child Eat School Breakfasts** |
| All Mothers | | 11.27% | 21.50% |
| Marital Status | | | |
| | Single Mothers | 22.18% | 34.30% |
| | Married Mothers | 7.19% | 16.80% |
| Mothers' Education | | | |
| | High School or Less | 18.07% | 31.40% |
| | Some College or Above | 6.41% | 16.10% |
| Mother's Race | | | |
| | Black | 21.27% | 40.90% |
| | White | 9.94% | 18.10% |
| Household Income | | | |
| | Income <= 185 FPL | 28.66% | 37.70% |
| | Income > 185 FPL | 3.48% | 12.20% |
| Central City Status | | | |
| | Central City | 16.64% | 30.53% |
| | Non-Central City | 9.78% | 14.31% |
| | | | |
| N | | 251,490 | 11,544 |

*Notes: This table summarizes the means of the School Breakfast Program participation using data from CPS Food Security Supplement during the period 1995-2012 and SIPP 2004 panel. We restrict the sample to mothers aged 19-50 and with at least one child aged 6-18.*
*Source: 1. CPS Food Security Supplement during the period 1995-2012. 2. SIPP 2004 Wave 3 panel data.*

# CHAPTER 2

## DO SMOKING BANS IMPROVE INFANT HEALTH?

## EVIDENCE FROM U.S. BIRTHS: 1995-2009[±]

Jia Gao

Department of Economics

University of New Hampshire

Reagan A. Baughman

Department of Economics

University of New Hampshire

[±] We are grateful to Karen Conway, Jennifer Trudeau, Hope Corman, Brandon Restrepo, Tianyan Hu, and anonymous referees for helpful comments. And we thank Eric Nesson for assistance with data. All errors are our own.

Corresponding author: Reagan Baughman, 370K Paul College of Business & Economics, University of New Hampshire, Durham NH 03824. (603) 862-0800. Reagan.Baughman@unh.edu

## 1. Introduction

Over the course of the past fifty years, following the publication of the first report on the health effects of smoking by the U.S. Surgeon General (USDHEW 1964), a multitude of health problems have been linked to tobacco use. One area in which smoking has been documented to have particularly serious adverse consequences is for infant health outcomes, including birth weight, length of gestation and birth defects (USDHHS 2004). Adverse birth outcomes, particularly low birth weight, have been linked to longer-term negative impacts on both child development and adult health. Treating infants with low birth weight also represents a substantial health care cost.

In response to these and other health problems linked to smoking, local, state and federal governments have developed policies designed to discourage consumption of cigarettes. Historically, the most widespread, and arguably successful, of these policies has been the cigarette tax. Although there is some disagreement in the literature as to the precise magnitude of the effect, pregnant women appear to be at least as sensitive to cigarette taxes as other demographic groups (Bradford 2003; Colman et al. 2003; Evans and Ringel 1999). During the past twenty years, state and local level smoking bans, in venues ranging from workplaces to bars and restaurants, have also become increasingly common.

Research on the effect of smoking bans on the general adult population has produced mixed results; while some studies have found significant reductions in smoking (e.g., Evans et al. 1999), other studies have found little to no effect (e,g., Bitler et al. 2010). However, relatively less is known about the impact of smoking bans on pregnant women and infants. Adams et al. (2012) show that private workplace bans increase third-trimester quit rates among pregnant women who smoked before becoming pregnant. Markowitz et al. (2013) find that state-level restrictions in restaurants and workplaces have very limited positive effects on birth weight in data from the Pregnancy Risk Assessment Monitoring System (PRAMS), which covers a subset of the US population. Briggs and Green (2012) look at the infant health effect of bans enacted before 2004; however, adoption rates for bans were low before 2005.

In this study, we use data from the U.S. Natality Detail File, which covers all of the births in the U.S. over the period 1995 to 2009 to test the impacts of smoking bans and cigarette taxes on birth weight and other measures of infant health. The availability of county-level geographic information in this dataset also allows for rich variation in policy, as we are able to measure the effects of ban adoption at both the state and local levels. Between 1995 and 2009, 27 states adopted new smoking bans and the average coverage level by at least one type of ban at the county level increased from 2 to 47 percent. Therefore, we are able to exploit a great deal of both cross-sectional and time variation in policy compared to existing studies. We also look at a rich set of measures of infant health, including very low birth weight, 5-minute APGAR scores and incidence of cleft lip/palate (a birth defect strongly linked to maternal smoking in the clinical literature).

Our results show that, in general, smoking bans are not associated with any changes in mean grams of birth weight or weeks of gestation. We also find that both restaurant/bar bans and workplaces bans are associated with small but meaningful *increases* of the probability of low birth weight and very low birth weight among infants born to women aged 14-24. The lack of health-improving effects implies that bans probably do not significantly reduce maternal smoking during pregnancy. We show that the negative infant health effects associated with smoking bans are found in babies born to mothers who reported not smoking at all during pregnancy – suggesting that increased exposure to second-hand smoke may be the mechanism for adverse effects.

## 2. Background

### 2.1 *Evidence on Smoking and Infant Health*

The detrimental effect of smoking on reproductive health is well established in the research literature. Smoking increases the risk of pregnancy complications such as placental abruption, premature rupture, as well as birth defects, and is also likely to lead to miscarriage, stillbirth, low birth weight, premature birth, fetal death and infant mortality (HHS 2004). Of these potential negative infant health outcomes, low birth weight (LBW, under 2,500 grams) and very low birth weight (VLBW, under 1,500 grams) stand out for several reasons. First, LBW and VLBW can lead to a higher risk of infant mortality and morbidity (McCormick 1985; Matthews 2001; Vogler and Kozlowski 2002). Second, the direct medical cost of LBW and VLBW in infants is extremely high and increases over time as medical technology improves (Almond et al. 2005). For babies who survive to age one, LBW and VLBW are also correlated with worse health outcomes later in life (McCormick et al. 1992; Corman and Chaikind 1998).

Additional adverse health outcomes for infants that previous studies have associated with smoking include preterm delivery (gestation < 37 weeks) (HHS 2004), low APGAR scores[1] (Kallen 2001; Garn et al. 1981) and birth defects such as limb reduction, clubfoot, oral cleft (cleft lip/palate), defect of the gastrointestinal system and cardiovascular defects (HHS 2004; Hackshaw et al. 2011; Honein et al. 2001). Of all of the birth defects mentioned above, smoking appears to have the largest effect on limb defects: the probability of delivering a baby with missing or

[1] A 5-minute APGAR test score is recorded for all infants five minutes after birth; the score is designed to provide information to doctors about how well the baby is doing outside mother's womb. The maximum possible score of 10 is made up of 0, 1 or 2 points in the following categories: color, heart rate, reflexes, muscle tone, respiration.

deformed limbs or a cleft lip is 25 percent higher for pregnant women who smoke compared to non-smoking pregnant women (Hackshaw et al. 2011).

### 2.2 *Public Policy and Smoking*

An early approach that significantly decreased maternal smoking during the 1990s in the United States was increasing cigarette tax rates (Evans and Ringel 1999; Ringel and Evans 2001; Bradford 2003). Clean indoor air ordinances or smoke-free laws are a more recent policy development; although local statutes to limit smoking in public places date to the 1970s, most of the original laws were not particularly stringent. During the late 1990s there was renewed interest in enacting smoke-free laws and by 2011, 79.4 percent of the U.S. population lived in an area with some type of ban on smoking in workplaces, restaurants, and/or bars (RWJF 2011).

Smoke-free laws/smoking bans are designed not only to improve the health of smokers, but also to protect the public from exposure to second-hand smoke. The majority of studies to date have found evidence that workplace restrictions or bans decrease cigarette smoking rates (Baile et al. 1991; Longo et al. 1996; Evans et al.1999; Adams et al. 2012). However, not all studies find a significant effect of bans on smoking behavior. Bitler et al. (2010) find no significant correlation between clean indoor air laws and reduction of smoking in workplaces, school areas, and restaurants, and they find that bar bans only decrease smoking for bartenders. The evidence of the impact of smoke-free policies on passive smoking (or second-hand smoke) is mixed, as well. Adda and Cornaglia (2010) find an unintended effect of smoking bans on children: bans in bars and restaurants increase second-hand smoke exposure by displacing smokers to private places, usually homes. However, Carpenter et al. (2010) do not find evidence of this kind of displacement effect when studying the public-place smoking restrictions in Canada.

If bans reduce smoking rates for pregnant women, it seems likely based upon the clinical literature that they should also improve infant health, especially by decreasing the incidence of LBW and of preterm delivery. However, the direction of any passive smoking effect is less clear. If, as Adda and Cornaglia (2010) suggest, a limitation on workplace smoking induces fathers to smoke more at home, pregnant mothers and infants may have worse health outcomes. Furthermore, smoking bans in restaurants and bars could lead more pregnant women to frequent these establishments, where they might consume alcohol or eat less healthy foods. These bans could also increase the number of pregnant women who choose to continue to work in bars and restaurants during pregnancy, which may also be detrimental to infant health.

### 2.3 *Evidence on Smoking Bans and Infant Health*

Recently, a handful of studies have emerged that look at the infant health effects of smoking bans put into effect outside of the United States (Kabir et al. 2009; Bharadwaj et al. 2014; Cox et al. 2013), but fewer studies consider the effect of smoking bans on infant health outcomes in the United States.

Amaral (2009) performs a case study for California to evaluate the impact of smoke-free ordinances in workplaces on infant health during the period 1988-2004. Her results indicate that workplace bans decrease average birth weight. Briggs and Green (2012) estimate the effect of workplace smoking bans of various stringencies on birth weight and gestation using data from Natality Detail File Data between 1989 and 2004. They estimate the effect of state-level ban adoption, while controlling for any pre-existing local bans as a proxy for voter demand for health, and find significant negative effects of the most stringent smoke-free laws in several cases. The study does not consider the effects of restaurant and/or bar bans. It is also based upon a sample period when adoption of bans was far less common than was in the late 2000s (as shown in Figure 1).

Markowitz (2008) finds that state-level bans in restaurants are associated with significantly lower rates of mortality from Sudden Infant Death Syndrome (SIDS) between 1973 and 2003, while smoking bans in workplaces has no significant effect on incidence of SIDS. Markowitz et al. (2013) look at the impact of cigarette taxes and smoking bans on infant birth weight and gestation. They find that workplace bans are not associated with any improvement in infant birth weight or gestation, while restaurant bans increase weeks of gestation for infants born to mothers aged 25-34. This study differs from ours in several ways. First, their data set (PRAMS) is not nationally representative; it ranges from 11 states in 1996 to 28 states in 2008. Additionally, the PRAMS is comprised of survey data attached to birth certificate records, and the response rate is only about 65% (70% in earlier years) overall, with lower rates for Black mothers, mothers having low birth weight infants, unmarried mothers and mothers with low education.[2] Our analysis is on the universe of births as recorded on birth certificates in the United States between 1995 and 2009. Finally, we are able to exploit greater variation in policy by measuring bans at the county rather than state level.

## 3. Data

### 3.1 *Infant Birth Data*

[2] The response rate and the availability of PRAMS data by year and by state is shown in CDC website: http://www.cdc.gov/prams/statesyearsdata.htm. Shulman et al. (2006) discuss the determine characteristics associated with the response rate of 2001 PRAMS.

The birth data come from the Natality Detail File, which contains information for all live births in United States. The data are taken directly from the birth certificate and provide information about infant birth date (year and month) and place (state and county)[3], demographic characteristics of mother and baby, and birth-related health outcomes, such as birth weight, gestational age, APGAR scores and complications during the birth. We use data from 1995 through 2009.

In order to construct a data set for analysis, we start with data on the full universe of births nationwide during this period. We drop approximately 26 percent of these observations in our baseline sample because they are for mothers who live in counties too small to be identified in the 1995-2004 public release data (population<250,000). [4] We then drop multiple births (approximately 3 percent of remaining sample) because the threshold for low birth weight is different for singleton and multiple births. We also exclude births to mothers below the age of 14 and above the age of 45 (approximately 3 percent of remaining sample) because of much higher risk of adverse outcomes at these ages. After dropping the certificates with missing information on birth weight and gestation (approximately 6 percent of remaining sample), we have just over 40 million observations.

The demographic control variables in our models will be mother's age, race, education, and marital status.[5] Unfortunately, birth certificates do not contain household income, which is likely to be an important factor that affects infant health. However, we do control for maternal

---

[3] Starting in 2005, geographic identifiers were only available in the restricted access version of the dataset, which we have used in this analysis.

[4] The missing of small counties (population < 250,000) before 2004 leads to an unbalanced panel for our study. We examine the robustness of results by dropping the small counties after 2004 in our specification checks.

[5] In the sample, approximately 1 percent of births have race missing values, and 0.5 percent of births have education missing values. In these cases, we construct dummy variables to indicate the missing education or race status.

education as a proxy for income. In addition, the birth order and the sex of the infant are also included in the model. Our main interest is the effect of smoking bans on infant health outcomes. For birth weight, we use both a continuous measure of birth weight in grams, as well as discrete variables to indicate whether the infants are LBW or VLBW. For gestation, we have both a continuous measure of gestation period in weeks and a discrete measure of low gestation (gestation<37 weeks). For 5-minute APGAR scores, we employ a single discrete variable to measure whether the infants have low APGAR scores (most commonly defined as below 7[6]). We also use a discrete variable to indicate the presence of cleft lip or palate.

Tables 1 and 2 describe the main variables from Natality Detail File and report their means and standard deviations for each maternal age group. The mean of birth weight is 3,238 grams for births to 14-24 year old mothers and 3,360 grams for births to mothers older than 25. Low birth weight is a relatively rare problem: LBW infants account for 6.3 percent of the full sample and VLBW infants account for just over 1 percent. The mean of gestation (between 38.72 and 38.57 weeks) changes very little by maternal age, with about 9 percent of infants having low gestation. Approximately 78 percent of birth certificate files contain a report on whether the mother smoked during pregnancy. By this measure, the prevalence of smoking appears to vary quite a bit by maternal age; mothers between 14 and 24 are about twice as likely to report smoking compared to mothers aged 35 to 45.

### 3.2 *Data on Smoking Bans*

[6] The most recent policy statement of the American Academy of Pediatrics and American College of Obstetricians and Gynecologist (2006) states that scores below 7 indicate the need for further medical attention and/or follow up.

Data on the local clean indoor air laws are from the American's Nonsmokers' Rights Foundation database. Our measures of smoking bans identify coverage of 100% smoke-free/clean indoor air laws for restaurants, bars and workplaces at the county level. The 100% smoke-free bans prohibit smoking with no exceptions. Because some of the bans are county-level and some are sub-county level (i.e. townships, municipalities), the variable used in our analysis is the fraction of a county's population in a given year that is covered by a given type of smoking ban[7]. For example, a ban variable with the value of one indicates all of the population within the county is fully covered by the smoking ban. If a ban variable is less than one (i.e., 0.50), it means that only a certain percentage (50%) of residents in that county are covered by the bans.

Over the period we study, the average fraction of county population covered by some type of smoking ban increased from 1.5 to 47 percent nationwide. Figure 1 shows the substantial growth in coverage by each type of bans. Among all three types of bans, bar bans were the least popular. In 1995, the coverage of bar bans was only approximately 0.1 percent at the county level across the country. Even in 2003, the coverage of bar bans was still limited to 3 percent. At the end of 2009, the coverage ratio had grown to 30 percent. Workplace bans were more common than bar bans, but still comparatively rare, in 1995, with an average 0.6 percent county-level coverage across the country. By 2009, the coverage ratio has almost risen to 40 percent. Following a similar trend, the coverage rate of restaurant bans went up from 1 percent in 1995 to 6 percent in 2003, and finally reached 40 percent in 2009.

## 4. Methods

[7] We are extremely grateful for Eric Nesson for sharing the data and programs that construct this measure with us.

Smoking bans have the potential to affect infant birth outcomes through both maternal cigarette consumption and the intake of second-hand smoke. Recognizing that there is more than one way in which policy may affect outcomes, we estimate a reduced form model to identify the *total* policy effect on birth outcomes. This total effect captures both direct effects arising from changes in maternal smoking behavior and indirect effects from changes in exposure to second-hand smoke or other factors. The model is:

$$\mathrm{Y}_{jisct} = X_{isct}\beta + \gamma_1 Ban_{ct} + \gamma_2 T_{st} + P_{sct}\lambda + \alpha_c + \alpha_t + \epsilon_{jisct} \quad (1)$$

where *Y* denotes the birth outcome *j* of infant *i* born in year *t* and whose mother is a resident in county *c* of state *s*, and *X* is a vector of infant and maternal demographic characteristics. The outcomes are: (1) birth weight in grams, (2) probability of low birth weight, (3) probability of very low birth weight, (4) weeks of gestation, (5) probability of low gestation, (6) probability of a low 5-minute APGAR score, and (7) probability of cleft lip/palate. All models are estimated by OLS with robust standard errors that are clustered at the state level.[8]

Bans vary by county and year; the coefficient of primary interest in the estimating equation is $\gamma_1$. Unobservable national time trends or cross-sectional state characteristics could bias our results if these factors were correlated with both infant birth outcomes and smoking policies, so we include both county ($\alpha_c$) and year ($\alpha_t$) fixed effects. Therefore, the coefficient $\gamma_1$ is identified exclusively by changes in county smoking policies over time. We also add county specific linear time trends to account for differences in trends across counties with bans and counties without

[8] For the discrete outcomes, as a specification check, probit models are also performed for a sub-sample of pregnant women living only in large counties (population > 250,000), which accounts for 90% of the full sample. The results are robust.

bans in our specification checks.[9] The significant amount of ban adoption that takes place during our study period ensures that this coefficient should be well identified, and the results are robust.

We control for state-level cigarette tax rates ($T$) using data from *The Tax Burden on Tobacco* (Orzechowski and Walker 2011). The cigarette tax variable is the tax on a pack of 20 cigarettes, adjusted for inflation to 2000 dollars. There are also several other state/county/year-varying policies that might affect infant birth outcomes; these controls are in the vector $P$. We include a simulated variable to account for changes in public health insurance eligibility. It is constructed the same way as Currie and Gruber (1996). Using a 1990 national sample, we calculate for each state and each year the percent of infant and children who would be eligible for Medicaid or State Child Health Insurance Programs (SCHIP). This variable varies only by legislative generosity within each state and over time, which does not capture the demographic characteristics of an actual state population that might affect infant health outcomes. We also control for the maximum Aid to Families with Dependent Children (AFDC) or Temporary Aid to Needy Families (TANF) benefit for a 3-person family in each state and year. County-level control variables in the model are: annual unemployment rate, per capita personal income and population.[10] We merge the data on bans and policy controls to the infant birth data by state/county of infant birth and quarter/year of conception, which is imputed using birth month, birth year and weeks of gestation.[11] The summary statistics of smoking bans, cigarette taxes and other state level controls are presented in Table 1.

---

[9] We are only able to run this model using a sub-sample restricted to pregnant women living in large counties (population>250,000).

[10] All monetary values are adjusted for inflation and expressed in 2000 dollars.

[11] The pregnant women are considered to be covered by the ban only when the implementation of the ban happens before or at the time of conception. If the ban is implemented during the pregnancy, the woman is considered not covered by the ban. Realizing that we code some women who are partially covered by the bans as not covered by the

## 5. Results

### 5.1 *Baseline Results*

We first estimate the effect of smoking bans on infant birth weight. The first three columns of Table 3 present results from reduced form OLS models for all maternal age groups when the dependent variable is birth weight in grams[12]. In the first row of Table 3, three types of bans are collapsed into a single variable denoting the existence of any type of ban. We create this *Any Ban* variable because there is a certain amount of collinearity in the adoption of different types of bans at the municipal and county level. However, the two major types of ban – restaurant/bar ban and workplace ban – also enter into the model separately to show each individual effect, and the results are presented in the second and third rows of the table.[13] Our results show that smoking bans do not have any significant effect on mean birth weight for any of the maternal age groups. Theoretically, the insignificant effect of *Any Ban* on birth weight could be because no types of bans have significant effects or because different types of bans have opposite effects that cancel one another out. The lack of significant results in the second and third rows of Table 3, in which restaurant/bar and workplace bans enter the model separately, supports the first explanation. The

bans, to improve the precision of estimation, we drop those women to examine the robustness of our results. We find the significance of our results does not change.

[12] All of our models estimated below include real state-level cigarette tax per pack, state and county-level control variables described in the previous section, and a full set of demographic control variables.

[13] For all the outcomes, we also estimate the models when restaurant/bar bans and workplace bans enter into the equation simultaneously. The results, similar to those shown in our baseline models, are not reported in the paper but are available upon request.

results for the other individual, state, and county level variables are presented in Appendix A. State and county level controls, including taxes, have very few significant effects.[14]

Previous studies have shown that anti-smoking policies can have non-linear effect on birth weight (e.g., Evan and Ringel 1999). If these policies only affect the most at-risk mothers, there might be policy effects on the incidence of LBW and VLBW, even though smoking policies do not affect birth weight at the mean. The results for probability of LBW and VLBW are in Columns 4-6 and 7-9 of Table 3, respectively. The results in the first row suggest that having any type of smoking ban is associated not with a decrease but with an *increase* of the probability of both LBW and VLBW for infants born to younger mothers (14-24). The estimates imply that a change from 0 percent to 100 percent any ban coverage would result in a 1.3 percent (0.001/7.49%) increase in the incidence of LBW and 3 percent (0.0004/1.32%) increase in the incidence of VLBW for infants born to women between 14 and 24. The results in the second row indicate that a change from 0 percent to 100 percent coverage by restaurant/bar bans would lead to a 3 percent increase in the incidence of VLBW infants born to mothers younger than 25. The results for workplace bans, presented in the third row, are similar to those in the first. While these percentage increases may appear small, they are likely to be economically significant in light of the enormous private and public costs associated with VLBW, in particular.

An infant health outcome that can be an important cause of low birth weight and also a potential health issue of its own is length of gestation. Our results for length of gestation in weeks are presented in the first three columns of Table 4. The overall gestation effect of smoking bans, as measured in the first row using the *Any Ban* variable, is insignificant. When the two types of

[14] Cigarette taxes do have significant, positive effects on birth weight for several vulnerable sub-groups, such as teenagers and younger black mothers. These results are available upon request.

smoking bans enter into the equation separately (see rows 2 and 3 of Table 4), we continue to find that neither type of bans are significantly associated with the length of gestation.

Noting that the estimates of the birth weight effects of smoking bans were different at the mean and the lower tail of the distribution, we also estimate linear probability models to examine whether or not smoking bans decrease the incidence of low gestation. The results in Columns 3-6 of Table 4 show no significant correlation between smoking policies and the probability of low gestation.

We next estimate the effect of smoking bans on the probability of a five-minute APGAR score being below 7. The results in the Columns 7-9 of Table 4 show that smoking bans do not have any significant effect on low APGAR scores. Approximately 97 percent of the infants in our sample have normal range APGAR scores, so it is difficult to detect any significant effect of bans.

Finally, we estimate a model to see whether there is a significant relationship between smoking bans and the incidence of cleft lip/palate. The results do not indicate that there is an effect for *Any Ban*, but there is a positive and significant coefficient on *Workplace Ban* in the second row of Table 4. Although the magnitude of this effect, a 0.02 percent increase, appears to be small, the underlying incidence of cleft palate (0.07 percent of all births) is so low that this is a meaningful one.

### 5.2 ***Results for Stratified Samples***

Previous research has documented significant differences in birth outcomes for black and white infants, so we stratify our sample by race and present the results for the outcomes LBW and

VLBW in the first four columns of Table 5.[15] We do not find any significant effect of smoking bans on infant health for either black or white infants. We also estimate the models for three vulnerable sub-groups: teenage mothers (14-19), black teenage mothers, and mothers with low levels of education. For the teenage mothers and black teenagers, the results (shown in Columns 5-8 of Table 5) indicate that both workplace bans and restaurants/bar bans are associated with significantly higher probability of VLBW, and the magnitude is larger than what we find in the baseline model. For the group of mothers with high school level of education or below, we fail to find any significant effects of smoking bans on the probability of having LBW or VLBW infants (Columns 9-10 of Table 5).

### 5.3 *Specification Checks*

We perform several specification checks to test the robustness of our baseline results.[16] First, anti-smoking sentiments may bias our results. If the passage of the clean indoor air laws in certain areas directly reflect more concern for health and distaste for smoking, then failing to control for differences in anti-smoking sentiment within state and over time will over-estimate the beneficial effects of smoking policies (DeCicca *et al*. 2008). In our case, the potential detrimental effects of smoking bans on infant health will be under-estimated. Following DeCicca *et al*. (2008), using data from Tobacco Use Supplement to the Current Population Survey (TUS-CPS) during the period 1995-2011, we include a direct measure of state-level anti-smoking sentiment as

[15] The results for the other birth outcomes are not shown in the paper, but available upon request. We do not find any significant results for the other outcomes among white or black infants.

[16] Because of the number of outcomes we have and models we estimate, we only present the results for the outcomes of LBW and VLBW in a few specification checks of interest in Table 6. Other results are available upon request.

additional explanatory variables in our baseline model.[17] This measure is based on responses to two questions on public attitudes towards smoking. In all TUS-CPS surveys during our study period, respondents were asked whether they think smoking should be allowed in bars and lounges, and whether they think smoking should be allowed at home. We create two variables – the percentage of people saying that smoking should not be allowed in bars and the percentage of people saying that smoking should not be allowed at home – to proxy for public attitudes towards smoking across states and over time. These variables are merged to the infant birth data based on state of residence and year of conception.

As shown in the first two columns of Table 6, the positive association between smoking bans and the incidence of LBW and VLBW infants born to mothers aged less than 24 still exists and is significant. When smoking bans are specified as *Any Ban* in the first row, the results indicate that a 0 percent to 100 percent change in coverage by any type of smoking ban would lead to an increase in the probability of having LBW and VLBW infants born to young mothers, and the size of the effects is similar to that we find in our baseline models. When each type of smoking bans enters into the equation separately, we continue to find a significant, positive effect for the outcome of VLBW and the magnitude doesn't change.

Second, we estimate models in which we control for pre-natal care during pregnancy. We do not include this control in our baseline models because it is likely to be endogenous (Rosenzweig and Schultz 1983). However, it may be an omitted variable if it affects smoking behavior during pregnancy and access varies by location and time. When we add a pre-natal care

[17] DeCicca *et al*. (2008) provide an in-depth discussion of the benefits of controlling for anti-smoking sentiment and present a way to measure this sentiment. Since our study period is longer than DeCicca *et al*. (2008), and TUS-CPS asked different questions after 2003, we are only able to use the replies to two questions (instead of nine questions in their paper) to measure the anti-smoking sentiment.

variable to our models, we still do not find any effect of smoking bans on mean birth weight or length of gestation, but the positive correlation between *Any Ban* and the rate of LBW and VLBW become insignificant (See Columns 3-4 of Table 6). However, pre-natal care is self-reported and incompletely reported in the birth certificate data, so this is not our preferred specification.

Third, the potential multi-collinearity between cigarette taxes and smoking bans may affect the significance of our results. To assess whether or not the time trend for cigarette tax increases is collinear with the trend in state adoption of smoking bans, we both informally observe whether there is a tax hike at the time of adoption of bans, and formally test the collinearity by calculating variance inflation factor (VIF). The VIF of workplace bans, bar/restaurant bans, and taxes are below 2.5, and the correlation between *Any Ban* and taxes is around 0.35, indicating no severe multi-collinearity. In addition, we replace the cigarette tax variable with a real cigarette price variable that includes taxes to test the robustness of results.[18] The results, in Columns 5-6 of Table 6, are not sensitive to this change.

Finally, we address the fact that in our 2005-2009 data, counties of all sizes are identified, while in the pre-2005 data, counties with populations below 250,000 are not identified, and so those observations with unidentified counties before 2005 are dropped in our baseline specifications. We estimate a set of models in which counties with populations below 250,000 were dropped in all years in our sample and report our results in Columns 7-8 of Table 6. Although the precision of the estimates falls slightly (some estimates are only significant at 10 percent level), the pattern of results is qualitatively similar to the baseline estimates. In the last two columns of Table 6, we report the results when including the county-specific linear time trends as additional

[18] Cigarette price data is from the Tax Burden on Tobacco (Orzechowski and Walker 2011). The price is a weighted average after-tax price for a pack of 20 cigarettes, varying by state and by year.

controls and using only observations in large counties (pop>250,000). As we can see, the smoking bans are still associated with higher rates of LBW and VLBW infants, although the significance level drops slightly.

## 6. The Potential Mechanisms for the Effects of Smoking Bans on Infant Health

Smoke-free laws are designed to discourage smoking and decrease the intake of second-hand smoke; in general, both would be expected to be clinically beneficial for pregnant women. At the same time, there are several possible explanations for the adverse health effects estimated in the previous section. The most likely explanation is increased exposure to second-hand smoke due to bans. This kind of change in the place of smoking is consistent with the findings of Adda and Cornaglia (2010). Alternately, pregnant women might be more likely to stay in the labor force after a workplace ban has been implemented or frequent bars or restaurants more often, all else equal.[19] Either of these scenarios would explain the adverse effect of bans on LBW and VLBW that we find in several specifications.

In this study of birth certificate records, which do not contain information on household structure or employment, we are limited in our ability to test all of these hypotheses. However, as one way of trying to disentangle the mechanism for these effects, we estimate a model in which the dependent variable is whether or not a mother reported smoking during pregnancy. As shown in Table 7, we find that smoking bans are not significantly associated with a reduction in smoking for pregnant women, which is consistent with our finding of no clinically beneficial effects of bans.

[19] Baum (2005) does not find any significant negative effects of working while pregnant, although a number of clinical studies document adverse effects of specific working conditions, including those requiring heavy lifting, exposure to chemicals, and other hazards.

To further test whether smoking status is a mediating factor in explaining the effects of bans, we add the maternal smoking during pregnancy variable to our baseline model. As expected, the results for the smoking bans do not change and the coefficient for smoking status is insignificant. All of these indicate that the effects of smoking bans on infant health do not work through the mechanism of decreasing maternal smoking.

This lack of an infant health effect through maternal smoking status still does not explain the negative effects we find on LBW and VLBW; however, these negative impacts of smoking bans could be due to redistribution of second-hand smoke from workplaces and restaurants into the home. To try to tease out this mechanism, we estimate specifications of models in which we stratify the sample between mothers who reported smoking during pregnancy and those who did not. The results shown in Table 8 suggest that the positive correlation between smoking bans and the incidence of LBW and VLBW exists *only* in non-smoker group. Although this is not a perfect test for effects of second-hand smoke exposure, the results suggest that negative infant health outcomes associated with smoking ban implementation work through indirect rather than direct channels.

Both of these analyses are limited attempts to understand the mechanism by which detrimental effects of smoking bans appear because the response rate to the smoking status question in the Natality Detail File is far from complete at 78 percent. This might explain why we fail to find any significant effects associated directly with smoking status. Additionally, birth certificate records do not provide information on the household of a new mother, including the kind of information on smoking behavior of all household member that would be needed to directly test the place-of-smoking hypothesis. Better understanding of the indirect health effects of smoking bans is an area for further research.

## 7. Conclusion

In conclusion, we have estimated the effect of smoking bans in restaurants/bars, and workplaces on infant birth weight, length of gestation and several other indicators of health at birth. Overall, we find no evidence that smoking bans are associated with improvements of infant health outcomes. On the contrary, the results suggest that the coverage of at least one type of smoking bans is associated with an increase in the incidence of LBW and VLBW among infants born to mothers younger than 25. A potential explanation for this negative effect is that pregnant women are exposed to more second-hand smoke after workplace bans are implemented because other household members smoke more in the home, an effect documented by Adda and Cornaglia (2010).

Despite years of policy attention, smoking and infant health outcomes such as low birth weight remain important public health concerns in most parts of the world. Our results suggest that bans on smoking in public places, which may have other amenity or adult health benefits, are not an effective way to reduce smoking during pregnancy or reduce smoking-attributable infant health problems. In fact, it is possible that bans may actually have a moderate detrimental effect on infant health by increasing the exposure of pregnant women through indirect channels, such as increased exposure to second-hand smoke, or changes in employment patterns for pregnant women. A better understanding of the mechanisms for these indirect effects remains an important goal for future work.

## LIST OF REFERENCES


Adams, E. Kathleen, Sara Markowitz, Viji Kannan, Patricia M. Dietz, Van T. Tong, and Ann M. Malarcher. 2012. "Reducing Prenatal Smoking: the Role of State Policies." *American Journal of Preventive Medicine*, 43(1):34-40.

Adda, Jerome, and Francesca Cornaglia. 2010. "The Effect of Bans and Taxes on Passive Smoking." *American Economic Journal: Applied Economics*, 2(1): 1-32.

Almond, Douglas, Kenneth Y. Chay, and David S. Lee. 2005. "The Costs of Low Birth Weight." *Quarterly Journal of Economics*, 20( 3): 1031-1083.

Amaral, Michelle. 2009. "The Effect of Local Smoking Ordinances on Fetal Development: Evidence from California". Working Paper.

American Academy of Pediatrics (AAP) and American College of Obstretricians and Gynecologists (ACOG). 2006. "Policy Statement: The APGAR Score." *Pediatrics,* 117(4): 1144-1147.

Bharadwaj, Prashant, Julian V. Johnsen, and Katrine V. Loken. 2014. "Smoking Bans, Maternal Smoking and Birth Outcomes". *Journal of Public Economics*, 115: 72-93.

Baile, Walter F., Michael Gibertini, Francis Ulschak, Sharon Snow-Antle, and Danette Hann. 1991. "Impact of a Hospital Smoking Ban: Changes in Tobacco Use and Employee Attitudes." *Addictive Behaviors*, 16(6): 419-426.

Baum, Charles. 2005. "The Effects of Employment While Pregnant on Health at Birth." *Economic Inquiry,* 43(2): 283-302.

Bitler, Marianne P., Carpenter Christopher, and Madeline Zavodny. 2010. "Effects of Venue-Specific State Clean Air Laws on Smoking-Related Outcomes." *Health Economics,* 19 (12): 1425-1440.

Bradford, David W.. 2003. "Pregnancy and the Demand for Cigarettes." *American Economic Review,* 93(5): 1752-1763.

Briggs, R.J., and Tiffany Green. 2012. "The Impact of Smoking Bans on Birth Weight: Is Less More?" Working Paper. Pennsylvania State University.

Carpenter, Christopher, Sabina Postolek, and Casey Warman. 2010. "Public-Place Smoking Laws and Exposure to Environmental Tobacco Smoke (ETS)." NBER Working Paper, No. 15849.

Colman, Greg, Michael Grossman, and Ted Joyce. 2003. "The Effect of Cigarette Taxes on Smoking Before, During and After Pregnancy." *Journal of Health Economics,* 22(6): 1053-1072.

Corman, Hope, and Stephen Chaikind. 1998. "The Effect of Low Birthweight on the School Performance and Behavior of School-aged Children." *Economics of Education Review*, 17(3): 307-316.

Cox, Bianca, Evelyne Martens, Benoit Nemery, Jacob Vangronsveld, and Tim S. Nawrot. 2013. "Impact of a Stepwise Introduction of Smoke-free Legislation on the Rate of Preterm Births: Analysis of Routinely Collected Birth Data." *BMJ: British Medical Journal,* 346: f441.

Currie, Janet, and Jonathan Gruber. 1996. "Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women." *Journal of Political Economy,* 104(6):1263-1296.

DeCicca, Philip, Donald Kenkel, Alan Mathios, Yoon-Jeong Shin, and Jae-Young Lim. 2008. "Youth Smoking, Cigarette Prices, and Anti-smoking Sentiment." *Health Economics,* 17(6): 733-749.

Evans, William N., and Jeanne S. Ringel. 1999. "Can Higher Cigarette Taxes Improve Birth Outcomes?" *Journal of Public Economics,* 72(1): 135-154.

Evans, William N., Matthew C. Farrelly, and Edward Montgomery. 1999. "Do Workplace Smoking Bans Reduce Smoking?" *The American Economic Review*, 9(4): 728-774.

Garn, Stanley M., Michael Johnston, Stephen A. Ridella, and Audrey S. Petzold. 1981. "Effect of Maternal Smoking on APGAR Scores." *American Journal of Diseases in Children,* 135: 503-506.

Hackshaw, Allan, Charles Rodeck, and Sadie Boniface. 2011. "Maternal Smoking in Pregnancy and Birth Defects: A Systematic Review Based on 173,687 Malformed Cases and 11.7 Million Controls." *Human Reproduction Update*, 17(5): 589-604.

Honein, M. A., L. J. Paulozzi, and M. L. Watkins, 2001. "Maternal Smoking and Birth Defects: Validity of Birth Certificate Data for Effect Estimation." *Public Health Report*, 116(4): 327-335.

Kabir, Z., V. Clarke, R. Conroy, E. McNamee, S. Daly, and L. Clancy. 2009. "Low Birthweight and Preterm Birth Rates 1 Year Before and After Irish Workplace Smoking Ban." *British Journal of Obstetrics and Gynaecology*, 116(13): 1782-1787.

Kallen, Karin. 2001. "The Impact of Maternal Smoking during Pregnancy on Delivery Outcome." *European Journal of Public Health,* 11(3): 329-333.

Longo, Daniel R., Ross C. Brownson, Jane C. Johnson, John E. Hewett, Robin L. Kruse, Thomas E. Novotny, and Robert A. Logan. 1996. "Hospital Smoking Bans and Employee Smoking Behavior." *Journal of the American Medical Association*, 275 (16): 1252-1307.

Markowitz, Sara. 2008. "The Effectiveness of Cigarette Regulations in Reducing Cases of Sudden Infant Death Syndrome." *Journal of Health Economics*, 27(1): 106-133.

Markowitz, Sara, E. Kathleen Adams, Patricia M. Dietz, Viji Kannan, and Van Tong. 2013. "Tobacco Control Policies, Birth Outcomes, and Maternal Human Capital." *Journal of Human Capital*, 7(2): 130-160.

McCormick, Marie C.. 1985. "The Contribution of Low Birth Weight to Infant Mortality and Childhood Morbidity." *New England Journal of Medicine*, 312(2): 82-90.

McCormick, Marie C., Jeanne B. Gunn, Kathryn W. Daniels, JoAnna Turner, and George J. Peckham. 1992. “The Health and Development Status of Very Low Birth Weight Children at School Age.” *Journal of American Medical Association*, 267(16): 2204-2208.

Orzechowski, W., and Walker, R. C.. 2011. “The Tax Burden on Tobacco: Historical Compilation 2010, vol. 37.” Arlington, Va: Orzechowski & Walker Consulting.

Ringel, Jeanne S., and William N. Evans. 2001. “Cigarette Taxes and Smoking During Pregnancy.” *American Journal of Public Health,* 91(11): 1851-1856.

Robert Wood Johnson Foundation (RWJF). 2011. “Interactive Tobacco Map.” http://www.rwjf.org/publichealth/product.jsp?id=56548.

Rosenzweig, Mark R. and Paul T. Schultz. 1983. “Estimating a Household Production Function; Heterogeneity, the Demand for Health Inputs, and Their Effects on Birth Weight.” *Journal of Political Economy*, 91(5): 723-746.

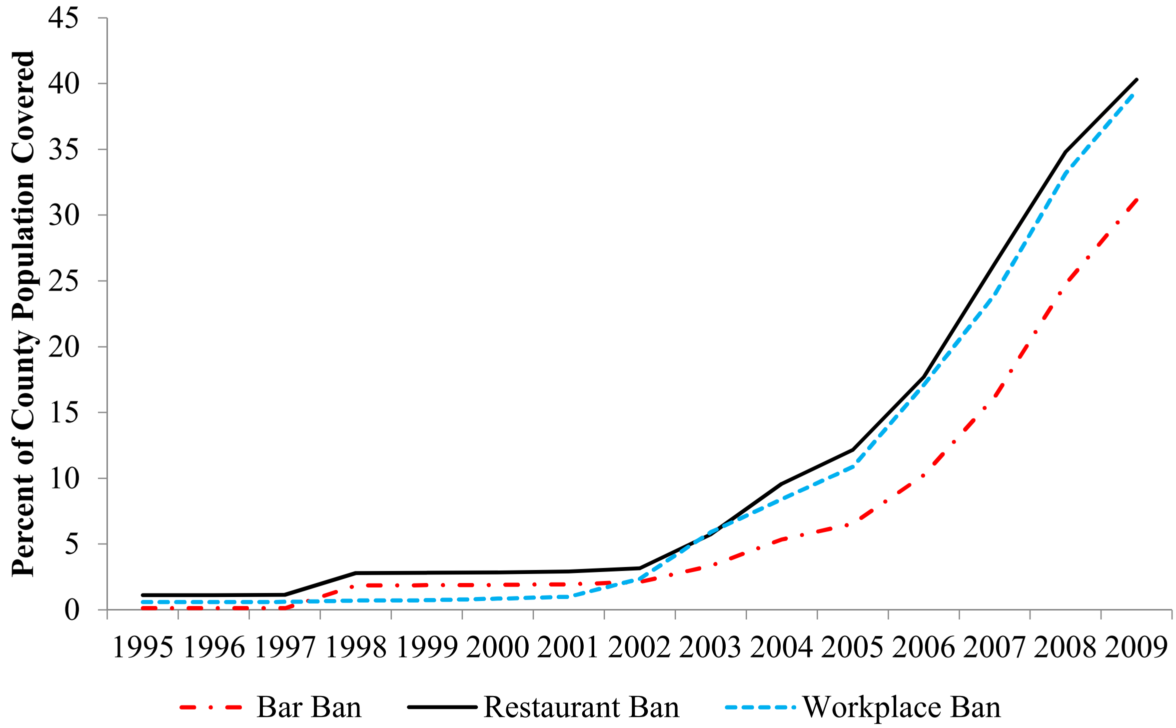
Shulman, Holly B., Brenda Colley Gilbert, and Amy Lansky. 2006. “The Pregnancy Risk Assessment Monitoring System (PRAMS): Current Methods and Evaluation of 2001 Response Rates.” *Public Health Reports*, 121(1): 74.

Vogler, George P., and Lynn T. Kozlowski. 2002. “Differential Influence of Maternal Smoking on Infant Birth Weight.” *The Journal of American Medical Association*, 287(2): 241-242.

U.S. Department of Health, Education and Welfare (USHEW). 1964. *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Washington: U.S. Department of Health, Education and Welfare, Public Health Service, Center for Disease Control. PHS Publication 1103.

U.S. Department of Health and Human Services. (USDHHS). 2004. *The Health Consequences of Smoking, A Report of the Surgeon General*. Public Health Service, Center for Disease Control and Prevention. National Center for Chronic Disease Prevention and Health Promotion Office on Smoking and Health, Office on Smoking and Health, Atlanta, Georgia.

**Figure 1. Percent of U.S. Population Covered by Smoking Bans at County Level, 1995-2009**



## Table 1. Descriptive Statistics: Outcome and Policy Variables

| Mother's Age | 14-24 | 25-34 | 35-45 | All |
|---|---|---|---|---|
| **Birthweight (grams)** | 3238.48 | 3358.17 | 3360.10 | 3316.95 |
| | (564.03) | (564.47) | (602.68) | (572.80) |
| **Low Birthweight (<2500g)** | 7.49% | 5.50% | 6.59% | 6.34% |
| | (0.26) | (0.23) | (0.25) | (0.24) |
| **Very Low Birthweight (<1500g)** | 1.32% | 1.01% | 1.31% | 1.16% |
| | (0.11) | (0.10) | (0.11) | (0.11) |
| **Gestation (weeks)** | 38.72 | 38.76 | 38.57 | 38.7165 |
| | (2.37) | (2.13) | (2.22) | (2.23) |
| **Low Gestation (<37 weeks)** | 10.06% | 8.40% | 9.91% | 9.19% |
| | (0.30) | (0.28) | (0.30) | (0.29) |
| **Infant Mortality** | 0.76% | 0.50% | 0.57% | 0.59% |
| | (0.09) | (0.07) | (0.08) | (0.08) |
| **Low Apgar Score (<7)** | 3.20% | 2.51% | 2.81% | 2.79% |
| | (0.18) | (0.16) | (0.17) | (0.16) |
| **Cleft Lip/Palate** | 0.07% | 0.07% | 0.07% | 0.07% |
| | (0.03) | (0.03) | (0.03) | (0.03) |
| **Mother Smoked** ± | 13.88% | 8.04% | 6.82% | 9.89% |
| | (0.35) | (0.27) | (0.25) | (0.30) |
| **Cigarette Tax$_{st}$** | $0.65 | $0.70 | $0.74 | $0.69 |
| | (0.45) | (0.47) | (0.49) | (0.47) |
| **Any Ban$_{ct}$** | 29.89% | 31.68% | 34.05% | 31.40% |
| | (0.44) | (0.45) | (0.46) | (0.45) |
| **AFDC/TANF$_{st}$** | 378.86 | 403.25 | 420.75 | 397.30 |
| | (156.06) | (154.54) | (152.23) | (155.43) |
| **Medicaid/SCHIP$_{st}$** | 44.44 | 45.08 | 46.24 | 45.03 |
| | (13.80) | (14.34) | (14.65) | (14.21) |
| **Unemployment$_{ct}$** | 5.48% | 5.24% | 5.17% | 5.32% |
| | (2.13) | (1.98) | (1.90) | (2.02) |
| **Per Capita Income$_{ct}$** | $30,000 | $32,000 | $33,900 | $31,600 |
| | (7723.08) | (8763.90) | (10000.00) | (8713.14) |
| **County Population$_{ct}$** | 1,330,000 | 1,360,000 | 1,450,000 | 1,360,000 |
| | (1,940,000) | (1,910,000) | (1,970,000) | (1,930,000) |
| **N** | 13,918,429 | 20,491,501 | 5,744,190 | 40,154,120 |

*Source: Infant birth outcomes are from U.S. Natality Detail File, 1995-2009. Tax data is from the Tax Burden on Tobacco (Orzechowski and Walker 2011). AFDC/TANF is collected from Green Book (U.S. House of Representatives). Medicaid/SCHIP is constructed by authors following Currie and Gruber (1996). Unemployment data is from Bureau of Labor Statistics. Income and population data are from Bureau of Economic Analysis. Standard deviations are in parentheses. ± Means and standard deviations for those who responded; response rates are 77% (14-24); 81% (25-34); 78% (35-44); 78% (all).*

**Table 2. Descriptive Statistics: Birth Certificate Demographic Variables**

| Mother's Age | 14-24 | 25-34 | 35-45 | All |
|---|---|---|---|---|
| **Mom Age** | 20.74 | 29.26 | 37.41 | 27.47 |
| | (2.40) | (2.78) | (2.27) | (6.18) |
| **Mom Hispanic** | 30.88% | 21.24% | 16.97% | 23.97% |
| | (0.46) | (0.41) | (0.38) | (0.43) |
| **Mom Black** | 23.22% | 12.91% | 11.79% | 16.32% |
| | (0.42) | (0.34) | (0.32) | (0.37) |
| **Mom Other Non-White** | 3.81% | 7.58% | 8.33% | 6.38% |
| | (0.19) | (0.26) | (0.28) | (0.24) |
| **Mom's Race Missing** | 0.75% | 0.96% | 1.16% | 0.91% |
| | (0.09) | (0.10) | (0.11) | (0.10) |
| **Mom Educ <8 yrs** | 7.39% | 5.35% | 5.23% | 6.04% |
| | (0.26) | (0.22) | (0.22) | (0.24) |
| **Mom Educ 8-11 yrs** | 29.58% | 8.43% | 5.30% | 15.31% |
| | (0.46) | (0.28) | (0.22) | (0.36) |
| **Mom Educ 13-15 yrs** | 19.33% | 24.85% | 21.94% | 22.52% |
| | (0.39) | (0.43) | (0.41) | (0.42) |
| **Mom Educ 16+yrs** | 3.82% | 35.53% | 46.16% | 26.06% |
| | (0.19) | (0.48) | (0.50) | (0.44) |
| **Mom Educ Missing** | 0.64% | 0.62% | 0.79% | 0.65% |
| | (0.08) | (0.08) | (0.09) | (0.08) |
| **Mom Married** | 35.12% | 76.44% | 82.94% | 63.05% |
| | (0.48) | (0.42) | (0.38) | (0.48) |
| **Prenatal Care Starts (Trimester)** | 1.31 | 1.16 | 1.14 | 1.21 |
| | (0.60) | (0.46) | (0.43) | (0.51) |
| **# Prenatal Care Visits** | 10.67 | 11.55 | 11.68 | 11.26 |
| | (4.13) | (3.82) | (3.97) | (3.98) |
| **1st child** | 57.90% | 34.12% | 23.35% | 40.82% |
| | (0.49) | (0.47) | (0.42) | (0.49) |
| **3rd child** | 9.91% | 18.86% | 22.45% | 16.27% |
| | (0.30) | (0.39) | (0.42) | (0.37) |
| **4th child** | 2.77% | 7.65% | 11.05% | 6.45% |
| | (0.16) | (0.27) | (0.31) | (0.25) |
| **5th child** | 0.90% | 4.76% | 10.61% | 4.26% |
| | (0.09) | (0.21) | (0.31) | (0.20) |
| **Birth Order Missing** | 0.58% | 0.53% | 0.59% | 0.56% |
| | (0.08) | (0.07) | (0.08) | (0.07) |
| **Male Child** | 51.20% | 51.22% | 51.17% | 51.20% |
| | (0.50) | (0.50) | (0.50) | (0.50) |

*Source: Author's tabulations from Natality Detail File, 1995-2009. Standard deviations are in parentheses.*

## Table 3. Baseline Ban Effects for Birth Weight, Low Birth Weight, and Very Low Birth Weight

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| **Mother Age** | **14-24** | **25-34** | **35-45** | **14-24** | **25-34** | **35-45** | **14-24** | **25-34** | **35-45** |
| **Outcome** | BW (g) | BW (g) | BW (g) | LBW | LBW | LBW | VLBW | VLBW | VLBW |
| **Any Ban** | -0.8954 | 0.4597 | 0.9906 | 0.0010* | 0.0003 | -0.0002 | 0.0004* | -0.0000 | -0.0000 |
| | (-0.57) | (0.29) | (0.57) | (2.59) | (0.75) | (-0.27) | (2.14) | (-0.24) | (-0.08) |
| **Restaurant/ Bar Ban** | 1.6462 | 1.8550 | 2.5041 | 0.0002 | 0.0002 | -0.0006 | 0.0004* | 0.0000 | -0.0002 |
| | (0.65) | (0.84) | (1.20) | (0.51) | (0.52) | (-0.87) | (2.54) | (0.02) | (-0.77) |
| **Workplace Ban** | -1.9251 | 0.4556 | 1.2995 | 0.0010 | 0.0005 | -0.0004 | 0.0004* | 0.0000 | -0.0002 |
| | (-1.27) | (0.26) | (0.70) | (1.84) | (1.10) | (-0.62) | (2.08) | (0.35) | (-0.92) |
| **N** | 13,918,429 | 20,491,501 | 5,744,190 | 13,918,429 | 20,491,501 | 5,744,190 | 13,918,429 | 20,491,501 | 5,744,190 |

*** Significant at the 1% level. * Significant at the 5% level. T-statistics are in parenthesis. All models also include: mother age, race, ethnicity, education, marital status; parity of birth; sex of child; county population, per capita income and unemployment rate; state AFDC/TANF benefit and Medicaid/SCHIP eligibility measure; cigarette tax; county and year fixed effects. Results for additional variables are presented in Appendix A.*

### Table 4. Ban Effects for Gestation, APGAR, and Cleft Lip, by Mother's Age

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Mother Age** | **14-24** | **25-34** | **35-45** | **14-24** | **25-34** | **35-45** | **14-24** | **25-34** | **35-45** | **14-24** | **25-34** | **35-45** |
| **Outcome** | GES | GES | GES | LGES | LGES | LGES | APGAR | APGAR | APGAR | Cleft Lip | Cleft Lip | Cleft Lip |
| **Any Ban** | 0.0033 | 0.0107 | 0.0119 | 0.0001 | -0.0007 | 0.0010 | 0.0034 | 0.0016 | 0.0004 | 0.0000 | -0.0000 | 0.0000 |
| | (0.21) | (0.63) | (0.79) | (0.09) | (-1.09) | (1.25) | (1.28) | (0.81) | (0.20) | (0.82) | (-1.23) | (0.99) |
| **Restaurant/ Bar Ban** | 0.0134 | 0.0170 | 0.0196 | -0.0008 | -0.0010 | -0.0015 | 0.0020 | 0.0006 | -0.0006 | 0.0000 | -0.0000 | 0.0000 |
| | (0.82) | (0.96) | (1.35) | (-0.82) | (-1.14) | (-1.66) | (0.72) | (0.28) | (-0.29) | (0.57) | (-1.60) | (0.45) |
| **Workplace Ban** | -0.0094 | -0.0002 | 0.0027 | 0.0007 | -0.0003 | -0.0012 | -0.0010 | -0.0013 | -0.0022 | 0.0000 | -0.0001 | 0.0001* |
| | (-0.64) | (-0.01) | (0.19) | (0.73) | (-0.38) | (-1.22) | (-0.45) | (-0.76) | (-1.29) | (0.80) | (-1.65) | (2.08) |
| **N** | 13,918,429 | 20,491,501 | 5,744,190 | 13,918,429 | 20,491,501 | 5,744,190 | 13,918,429 | 20,491,501 | 5,744,190 | 13,918,429 | 20,491,501 | 5,744,190 |

*** Significant at the 1% level. * Significant at the 5% level. T-statistics are in parenthesis. All models also include: mother age, race, ethnicity, education, marital status; parity of birth; sex of child; county population, per capita income and unemployment rate; state AFDC/TANF benefit and Medicaid/SCHIP eligibility measure; cigarette tax; county and year fixed effects.*

**Table 5. Ban Effects Stratified by Race, Age and Education**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| | **White** | | **Black** | | **Teen** | | **Black Teen** | | **Low Education** | |
| **Outcome** | LBW | VLBW | LBW | VLBW | LBW | VLBW | LBW | VLBW | LBW | VLBW |
| **Any Ban** | -0.0008<br>(-0.87) | -0.0004<br>(-1.24) | -0.0019<br>(-0.91) | -0.0008<br>(-0.92) | 0.0010<br>(1.39) | 0.0008*<br>(2.17) | 0.0015<br>(1.02) | 0.0017*<br>(2.15) | 0.0005<br>(1.19) | 0.0000<br>(0.05) |
| **Restaurant \Bar Ban** | -0.0005<br>(-0.53) | -0.0002<br>(-0.53) | -0.0009<br>(-0.49) | 0.0002<br>(0.21) | 0.0004<br>(0.51) | 0.0008*<br>(2.28) | 0.0014<br>(0.93) | 0.0020*<br>(2.62) | -0.0000<br>(-0.07) | 0.0000<br>(0.17) |
| **Workplace Ban** | 0.0004<br>(0.43) | 0.0001<br>(0.24) | 0.0001<br>(-0.04) | 0.0002<br>(0.21) | 0.0014<br>(1.80) | 0.0009*<br>(2.48) | 0.0016<br>(1.10) | 0.0019*<br>(2.09) | 0.0005<br>(0.69) | 0.0000<br>(0.13) |
| **N** | 20,307,031 | 20,307,031 | 6,109,409 | 6,109,409 | 4,293,465 | 4,293,465 | 1,197,561 | 1,197,561 | 20,387,713 | 20,387,713 |

*** Significant at the 1% level. * Significant at the 5% level. T-statistics are in parenthesis. All models also include: mother age, ethnicity, education, marital status; parity of birth; sex of child; county population, per capita income and unemployment rate; state AFDC/TANF benefit and Medicaid/SCHIP eligibility measure; cigarette tax; county and year fixed effects*

## Table 6. Specification Checks

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Specification 1: Control for Smoking Sentiment** | | **Specification 2: Control for Prenatal Care** | | **Specification 3: Control for Cigarette Price instead of Tax** | | **Specification 4: Restrict Sample to Large Counties** | | **Specification 5: With County-Specific Trend** | |
| **Outcome** | LBW | VLBW | LBW | VLBW | LBW | VLBW | LBW | VLBW | LBW | VLBW |
| **Mother Age** | 14-24 | 14-24 | 14-24 | 14-24 | 14-24 | 14-24 | 14-24 | 14-24 | 14-24 | 14-24 |
| **Any Ban** | 0.0012* | 0.0005* | 0.0003 | 0.0000 | 0.0010* | 0.0004$^{\pm}$ | 0.0011* | 0.0004$^{\pm}$ | 0.0010* | 0.0004$^{\pm}$ |
| | (2.48) | (2.10) | (0.32) | (0.12) | (2.07) | (1.87) | (2.26) | (1.70) | (2.30) | (1.77) |
| **Restaurant \Bar Ban** | 0.0005 | 0.0005** | 0.0003 | 0.0004 | 0.0002 | 0.0004* | 0.0002 | 0.0004* | 0.0019 | 0.0004* |
| | (1.16) | (2.75) | (0.27) | (0.82) | (0.43) | (2.38) | (0.40) | (2.29) | (0.45) | (2.17) |
| **Workplace Ban** | 0.0010$^{\pm}$ | 0.0004* | 0.0007 | 0.0001 | 0.0010 | 0.0004$^{\pm}$ | 0.0013 | 0.0004$^{\pm}$ | 0.0011 | 0.0004$^{\pm}$ |
| | (1.71) | (2.08) | (0.53) | (0.28) | (1.67) | (1.90) | (1.84) | (1.82) | (1.63) | (1.85) |
| **N** | 13,918,429 | 13,918,429 | 13,224,197 | 13,224,197 | 13,918,429 | 13,918,429 | 10,837,912 | 10,837,912 | 10,837,912 | 10,837,912 |

*** Significant at the 1% level. * Significant at the 5% level. $^{\pm}$ Significant at the 10% level. T-statistics are in parenthesis. All models also include: mother age, ethnicity, education, marital status; parity of birth; sex of child; county population, per capita income and unemployment rate; state AFDC/TANF benefit and Medicaid/SCHIP eligibility measure; county and year fixed effects.*

**Table 7. Ban Effects on Whether Mother Smoked During Pregnancy, by Age**

| Outcome | (1)<br>14-24<br>Smoked | (2)<br>25-34<br>Smoked | (3)<br>35-45<br>Smoked |
|---|---|---|---|
| **Any Ban** | -0.0003 | 0.0001 | 0.0021 |
| | (-0.10) | (0.03) | (1.08) |
| **Restaurant \Bar Ban** | 0.0012 | 0.0002 | 0.0037 |
| | (0.28) | (0.06) | (1.73) |
| **Workplace Ban** | -0.0024 | -0.0005 | -0.0015 |
| | (-0.70) | (-0.24) | (-0.76) |
| **N** | 13,918,429 | 20,491,501 | 5,744,190 |

*** Significant at the 1%* level*. * Significant at the 5% level. T-statistics are in parenthesis. All models also include: mother age, ethnicity, education, marital status; parity of birth; sex of child; county population, per capita income and unemployment rate; state AFDC/TANF benefit and Medicaid/SCHIP eligibility measure; county and year fixed effects.*

**Table 8. Ban Effects Stratified by Mother's Reported Smoking Status**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | **14-24** Non-smokers | **25-34** Non-smokers | **35-46** Non-smokers | **14-24** Smokers | **25-34** Smokers | **35-46** Smokers |
| **Outcome** | BW (g) | BW (g) | BW (g) | BW (g) | BW (g) | BW (g) |
| **Any Ban** | -2.1308 | -0.4676 | -0.4388 | -0.6335 | 0.0898 | -2.2095 |
| | (-1.29) | (-0.13) | (-0.26) | (-0.15) | (0.04) | (-0.36) |
| **Outcome** | LBW | LBW | LBW | LBW | LBW | LBW |
| **Any Ban** | 0.0020** | 0.0009* | 0.0006 | 0.0014 | 0.0001 | 0.0017 |
| | (3.24) | (2.25) | (1.88) | (0.83) | (0.10) | (0.52) |
| **Outcome** | VLBW | VLBW | VLBW | VLBW | VLBW | VLBW |
| **Any Ban** | 0.0006*** | 0.0003 | 0.0000 | 0.0004 | -0.0001 | 0.0003 |
| | (3.59) | (0.59) | (0.01) | (0.78) | (-0.35) | (0.24) |
| **N** | 10,102,037 | 15,009,406 | 4,382,778 | 1,261,607 | 1,317,871 | 305,811 |

****Significant at the 0.1% level. ** Significant at the 1% level. * Significant at the 5% level. T-statistics are in parenthesis. All models also include: mother age, race, ethnicity, education, marital status; parity of birth; sex of child; county population, per capita income and unemployment rate; state AFDC/TANF benefit and Medicaid/SCHIP eligibility measure; cigarette tax; county and year fixed effects.*

**Appendix Table 1. Coefficients for Individual Control Variables, all Policy and County Characteristic Variables: Birth Weight Models**

| Mother's Age | 14-24 | 25-34 | 35-45 |
|---|---|---|---|
| **Outcome** | BW (g) | BW (g) | BW (g) |
| **Any Ban$_{ct}$** | -0.8954 | 0.4597 | 0.9906 |
| | (-0.57) | (0.29) | (0.57) |
| **Mom Age** | 8.6689*** | -0.2447 | -8.1603*** |
| | (14.84) | (-1.63) | (-19.59) |
| **Mom Hispanic** | -27.8150*** | -19.2341*** | -28.9358*** |
| | (-5.86) | (-4.18) | (-5.56) |
| **Mom Black** | -190.2260*** | -210.9928*** | -238.2339*** |
| | (-40.17) | (-32.92) | (-27.54) |
| **Mom Other** | -120.2343*** | -175.4238*** | -164.0369*** |
| **Non-White** | (-6.24) | (-21.57) | (-30.22) |
| **Mom Edu<8** | -20.5404*** | 17.4632** | 31.0553*** |
| | (-4.23) | (3.27) | (6.34) |
| **Mom Edu 8-11** | -31.3293*** | -40.3134*** | -47.0533*** |
| | (-10.00) | (-5.51) | (-4.57) |
| **Mom Edu 13-15** | 29.8958*** | 34.6638*** | 43.2286*** |
| | (11.71) | (13.95) | (15.44) |
| **Mom Edu 16+** | 46.1847*** | 67.0804*** | 90.3018*** |
| | (11.82) | (15.50) | (18.57) |
| **Mom Married** | 40.4126*** | 75.1733*** | 106.5997*** |
| | (13.53) | (14.55) | (13.79) |
| **1$^{st}$ Child** | -43.3547*** | -99.0247*** | -121.2960*** |
| | (-15.46) | (-56.74) | (-47.48) |
| **3$^{rd}$ Child** | -19.2232*** | 12.0210*** | 32.4250*** |
| | (-10.20) | (7.72) | (37.68) |
| **4$^{th}$ Child** | -48.2554*** | 3.6987 | 49.6900*** |
| | (-17.35) | (1.05) | (32.74) |
| **5$^{th}$ Child** | -80.2204*** | -18.6504*** | 74.3996*** |
| | (-29.05) | (-3.54) | (17.92) |
| **Male** | 103.5531*** | 115.0347*** | 115.0540*** |
| | (84.87) | (93.89) | (70.87) |
| **Cigarette Tax$_{st}$** | 3.3249 | 0.2372 | -0.1657 |
| | (1.27) | (0.11) | (-0.06) |
| **AFDC/TANF$_{st}$** | -0.1102* | -0.0477 | -0.0390 |
| | (-2.51) | (-1.31) | (-1.26) |
| **Medicaid/SCHIP$_{st}$** | 0.0270 | -0.0395 | -0.1340 |
| | (0.29) | (-0.63) | (-1.67) |
| **Unemployment$_{ct}$** | 0.1232 | 0.2970 | 0.3958 |
| | (0.17) | (0.37) | (0.62) |
| **Per Capita Inc$_{ct}$** | 0.0003 | 0.0002 | 0.0001 |
| | (1.01) | (0.46) | (0.37) |
| **Population$_{ct}$** | -0.0000 | -0.0000 | -0.0000 |
| | (-0.79) | (-1.51) | (-1.70) |
| **Observations** | 13,918,423 | 20,491,502 | 5,744,195 |

**** Significant at the 0.1% level. ** Significant at the 1% level. * Significant at the 5% level. T-statistics are in parenthesis. All models also include mom education missing, mom race missing, and birth order missing variables, but the coefficients are not reported here. The omitted categories are: white mother, high school education, and second live birth.*

# CHAPTER 3

## LANGUAGE ENVIRONMENT AT HOME AND ACADEMIC ACHIEVEMENT DURING EARLY CHILDHOOD

Jia Gao

Department of Economics

University of New Hampshire

Le Wang[±]

Department of Economics, Finance, and Legal Studies

University of Alabama & IZA

---

[±] Corresponding author: Le Wang, Department of Economics, Finance, and Legal Studies, University of Alabama, Tuscaloosa, AL 35473. Email: Le.Wang.Econ@gmail.com; Phone: 205 348 8967; Fax: 205 348 0590.

## 1. Introduction

Over the past two decades, there has been rapid growth in the immigrant population in the United States; the number of immigrants has increased from 19.8 million to 38.1 million during the period 1990 to 2007 (Fortuny and Chaudry 2009). This dramatic increase has fueled a rise in the share of children with immigrant parents. In 2007, children of immigrants represented more than one in five American children (Fortuny and Chaudry 2009). The gap in educational achievement between the children of immigrants and their native peers is well documented (Schnepf 2007; Dustmann and Glitz 2011). This gap appears to start at a very early age and persist into adulthood. For example, data from the National Assessment of Educational Progress in 2008 shows that the average reading score gap between white students and Hispanic students was roughly 10 percent for students aged 9-, 13- and 17-years-old (Rampney et al. 2009). A relatively smaller gap was also found for math scores: 7 percent for 9-year-old students, and 8 percent for 13- and 17-year-old students. These figures raise serious concerns about immigrants' ability to assimilate into the US economy because this education is an important precondition for them to integrate. (Schnepf 2007). From a policy perspective, it is also necessary for governments to support a productive and highly qualified workforce that will meet the challenges of an aging population (Colding et al. 2009).

The majority of children of immigrants are exposed to a multi-lingual environment, and many of them report that their primary language spoken at home is not English. For example, using Poverty and Social Impact Analysis (PSIA) data, Dustmann et al. (2011) find that 62.3 percent of children of immigrants are language minority students (or students who speak a foreign language at home and have English as an additional language). Given that home learning environment affects student academic achievement, it is not surprising that the literature finds that among many factors (such as parental socioeconomic status and school characteristics), language environment at home—specifically, whether the primary language spoken at home is English—is one of the most important factors in explaining the test score gap (for example, Schnepf 2007). However, in most previous studies, language environment at home is usually modeled only as a control variable. Its importance is measured by how much the immigrant status effect on educational achievement shrinks after controlling for languages spoken at home. Although these studies imply that language environment at home can play a significant role in promoting students' academic success, the estimated relationship between language spoken at home and school performance has not been of primary interest. More importantly, the issue of endogeneity is often ignored, and this may bias the estimates. For example, students speaking only English at home may have parents who care more about child education and are able to provide more help to their children in academics. Without controlling for these unobserved characteristics, the OLS estimates are biased.

Our paper is among the first to attempt to fill this gap. Specifically, we focus on estimating the *causal* effect of language environment at home, measured by whether one speaks a language other than English at home, on students' test scores during early childhood. The data used in this paper comes from Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K). Using the confidential ECLS-K data that allow us to match zip codes to county codes in the American Community Survey, we are able to exploit the variation in local language environment and augment this external instrument with a novel internal instrumental variables (IV) approach developed in Lewbel (2012), to isolate the causal effects. We reach several conclusions. First, we find that speaking a language other than English at home has a sizable, negative impact on reading test scores in both third and fifth grades, but has no effect on math scores in either grade. Second, there is no evidence that speaking a language other than English at home has any effect on the growth rate of test scores from the third to the fifth grade, regardless of the subject. This suggests that any significant negative effect we find on reading test scores in the fifth grade is a result of the gap in the third grade. A non-English-language environment at home does not increase or decrease the test score gap between English users and non-English users from the third grade to the fifth grade. Third, the pattern in the effects of language environment remains unchanged whether the foreign language is Spanish or another language. Finally, our results also indicate that the effects of language environment at home differ by gender. Speaking a language other than English at home decreases the scores of girls on both reading and math tests, while for boys, this detrimental effect only exists for reading scores.

The literature has generally found that cognitive and non-cognitive skills developed in early childhood can have a long-lasting impact on future educational and labor market outcomes (Keane and Wolpin 1997; Keane and Wolpin 2001; Cameron and Heckman 2001; Heckman 2008). Therefore, our results raise concerns about the long-term well-being of the children of immigrants whose primary language at home is not English, and may suggest the need for potential policy interventions.

The remainder of the paper is organized as follows. Section 2 reviews the literature and discusses the potential mechanisms through which language environment at home could affect test scores. Section 3 presents the empirical methods. Section 4 describes the data. Section 5 reports the main results. Section 6 presents the results across different sub-groups. Finally, Section 7 concludes.

## 2. Literature Review

In this section we review two streams of the literature on the relationship between bilingualism and school performance. We first provide a review of the competing theories that have been put forth to explain an association between language environment at home and children's academic achievement. We categorize them by whether or not the association is a direct one. We then review some of the related empirical literature.

## 2.1 *Links between Language Spoken at Home and Academic Achievement*

### *2.1.1 Direct Effects*

Language spoken at home may affect school performance directly by influencing the fundamental aspects of cognitive and neural development of children. A substantial body of research in linguistics and sociology has evaluated the effect of foreign language study or bilingualism on child cognitive development. There are two opposing views on the relationship between bilingualism and cognitive development: (1) the subtractive view, and (2) the additive view.

The subtractive view emphasizes the negative effect of bilingualism and second language learning on cognitive development of children. This view was particularly popular in studies before 1970s. Researchers have argued that second language users are less effective in processing two languages at the same time (August and Hakuta 1997; Diaz 1983). They argue that human beings naturally begin with one language; therefore, a child has an extra cost of speaking an additional language because she usually needs to transfer "code" from the first language to the second language. During the process of transferring, the child may confuse herself both linguistically and cognitively, especially a child who has just begun to learn a second language. Additionally, speakers of multiple languages need separate systems to process different languages, so that using one language interferes with using the other (for reviews, see Cook 1997; Bialystok et al. 2012). More recently, researchers have provided the evidence that monolinguals generally perform better than bilinguals on verbal tasks, such as identifying and naming pictures more rapidly, because bilinguals master a smaller vocabulary in each language than monolinguals (Bialystok 2009; Pelham and Abrams 2014). This finding suggests a potential negative effect of bilingualism on students' oral and written language development.

The additive view, on the other hand, emphasizes the positive effect of bilingualism on cognitive development. It holds that knowing a second language improves rather than shrinks cognitive capability and leads to higher levels of metalinguistic awareness (Hakuta 1986; Bialystok 1999; Bialystok and Martin 2004). By learning another language, a student can improve writing and speaking skills, and even sharpen the awareness of the first language because more attention is paid to pronunciation and grammar, which may be hardly noticed when using only the native language (Bialystok 2013; Homel et al. 2014). Additionally, using multiple languages improves divergent thinking, which values flexibility, originality and fluency, increases creativity, and facilitates analogical reasoning (Bialystok 2007; Yoshida 2008; Abutalebi and Clahsen 2015).

### *2.1.2 Indirect Effects*

English skill is another way through which languages spoken at home could indirectly affect students' academic performance. Some immigrant parents speak English at home with the purpose of helping their children to learn English. The economics literature has found that speaking English at home improves language proficiency within immigrant families (for example, see Chiswick et al. 2005). The psychobiology literature generally finds that younger children learn language more easily than adolescents. It is widely believed that there is a "critical period" age range in which learning English happens naturally and can be considered as a natural ability for children (Bleakley and Chin 2004). While there is no general consensus on the critical period age range, most studies find supportive evidence for the critical period hypothesis, with a range between age 5 and age 15 (Chiswick and Miller 2008). In our sample, all children fall into this range. Thus, it is possible that learning English takes place within households. If that is the case, children of immigrants would benefit from speaking English at home, and even may perform better in cognitive tests. Although this indirect effect may exist, it is important to be aware that all language minority students in our sample have passed an English screening test, and so schools believe that their English skills do not affect the understanding of the cognitive tests. Therefore, the effect of languages used at home on test scores found in our sample captures more than merely English skills.

In addition to this language skill effect, language environment at home may affect non-cognitive development, and eventually influence children's academic performance. For immigrant families, using the native language of parents at home may have beneficial effects for children by promoting verbal interaction and complexity of communications between immigrant parents and their young children (Greenspan 1997).

The discussion here is not intended to be exhaustive. Instead, we want to highlight the possibly important role of language environment at home in determining children's academic achievement. This in turn motivates us to provide a first attempt to empirically examine this issue.

### 2.2 *Related Empirical Literature*

Our paper is closely related to two strands of the empirical literature. First, a growing literature has attempted to estimate the relationship between bilingualism and test scores among young adults. Using data on students from Brigham Young University (BYU) who were sent to

Mormon missions randomly and had to learn the foreign language of their mission's location, Pope (2008) compares the difference in GPA for students who served foreign-speaking missions relative to students who served English-speaking missions. Pope fails to find evidence that students who were assigned to speak a foreign language perform better than students who were assigned to speak English. While this study, based on randomized experiments, provides some convincing results, the results cannot necessarily be generalized to younger children, who are the focus of this paper.

Another strand of the literature focuses on the relationship between language proficiency and academic achievement among immigrants (Dustmann et al. 2010; Dustmann and Trentini 2008; Schnepf 2007). Most immigrants face a language obstacle that could potentially impact their academic performance at school. Dustmann et al. (2010) find that language is the single most important factor in explaining why the achievement gap between ethnic minority pupils and white British pupils decreases over time, using national pupil data from the UK. In the United States, one of the hotly debated issues is whether to provide language minority students with a bilingual education system or to immerse them in a purely English environment. Several studies have examined the effectiveness of using native language instruction for limited English proficient students, but they reach different conclusions. Chin et al. (2013) look at the impact of bilingual education programs in Texas on student development. They find that bilingual education programs do not improve test scores for limited English proficiency students. Slavin et al. (2011) and Matsudaira (2005) also provide the evidence that native language instruction neither helps language minority students improve English skills nor enhances their cognitive development. However, Greene (1998) conducts a meta-analysis by summarizing 11 studies, and finds that using a child's native language to teach him could help him to learn.

## 3. Empirical Methods

Before estimating the impact of speaking a language other than English at home on test scores, we first discuss a framework of potential outcomes to motivate our estimators. To begin with, let $Y_1$ and $Y_0$ denote two test scores to be compared. Specifically, $Y_1$ represents the test score that a student *i* would get if she spoke a language other than English at home, while $Y_0$ is the test score that student *i* would receive if she spoke English at home. In practice, only the data for one of the two test scores is collected; thus, one is the true score, and the other is a hypothetical score. The structural relationship between potential outcomes and the choice of language spoken at home is as follows:

$$Y_{i1} = \alpha + \delta_i + \mu_1(X_i) + u_{i1} \tag{1}$$

$$Y_{i0} = \alpha + \mu_0(X_i) + u_{i0} \tag{2}$$

$$D_i = \begin{cases} 1 & X_i\lambda + \epsilon_i > 0 \\ 0 & other\ wise \end{cases} \tag{3}$$

where $X_i$ is a vector of exogenous observable characteristics; $\delta_i$ represents the effect of speaking a language other than English at home on test scores; $\mu_{i1}$ and $\mu_{i0}$ capture the unobservable determinants of test scores; and $D_i$ is a dummy variable indicating whether the language spoken at home is English, and equal to 1 if speaking a language other than English and zero otherwise. The observed test scores for student *i* could be expressed as $Y_i = D_iY_{i1} + (1 - D_i)Y_{i0}$.

In this set-up, the effect of speaking a non-English language at home is given by $Y_{i1} - Y_{i0}$, which varies from person to person. Here, we focus on the average treatment effect (ATE), which is $E[Y_{i1} - Y_{i0}]$. If we assume: (1) constant effects of observables across families speaking English and families not speaking English at home, $\mu_1(X_i) = \mu_0(X_i) = X_i\beta$; (2) common unobservables across the two different home environments, $u_{i1} = u_{i0} = u_i$; and (3) constant home language effect across individuals, $\delta_i = \delta$, the ATE simplifies to:

$$
\begin{aligned}
E[Y_{i1} - Y_{i0}] &= E[(\alpha + \delta_i + \mu_1(X_i) + u_{i1}) - (\alpha + \mu_0(X_i) + u_{io})] \\
&= E[(\alpha + \delta + X_i\beta + u_i) - (\alpha + X_i\beta + u_i)] \\
&= \delta \qquad (4)
\end{aligned}
$$

Under these assumptions, the observed test score $Y_i$ can be estimated using the following regression model:

$$
\begin{aligned}
Y_i &= D_iY_{i1} + (1 - D_i)Y_{i0} \\
&= D_i(\alpha + \delta_i + \mu_1(X_i) + u_{i1}) + (1 - D_i)(\alpha + \mu_0(X_i) + u_{i0}) \\
&= D_i(\alpha + \delta_i + X_i\beta + u_i) + (1 - D_i)(\alpha + X_i\beta + u_i) \\
&= \alpha + \delta D_i + X_i\beta + u_i \qquad (5)
\end{aligned}
$$

### 3.1 *OLS Model*

The most common estimator used to identify the effect of language environment at home on child school performance is the Ordinary Least Squares estimator, which is given by $\delta_i$ in Equation (5). The estimate of $\delta_i$ will only be unbiased if there are no other unobservable characteristics correlated with both children's test scores and the choice of language spoken at home. In other words, in Equation (5), if conditional on $X_i$, $D_i$ is independent of $u_i$ or $E[u_i|D_i; X_i] = 0$, then we can consistently estimate the ATE ($\delta_i$) by OLS.

Consistent and unbiased OLS estimation assumes that the choice of language spoken at home is independent of other unobservable factors within the households. However, in the real world, the decision about whether or not to speak English at home is correlated with a number of factors, such as parents' English proficiency, socioeconomic status, and preference for keeping origin languages. Some of these factors are also correlated with children's school performance, but are very difficult to observe and measure. For instance, parents' English proficiency is an important factor in determining whether to speak English at home or not, and it will also affect how parents educate their children (such as whether they read stories in English to children), which would influence child test scores. Another concern is that children's English skills, which are correlated with their test scores, may affect parents' choice of language used at home. For example, children's poor school performance due to limited English skills could push parents to speak English at home so as to improve children's school performance. Therefore, it is very likely that in Equation (5) $E[u_i|D_i; X_i] \neq 0$; that is, there is an endogeneity issue, and the OLS estimator cannot deliver a consistent estimate. A typical strategy to address the endogeneity is to employ instruments.[1] In the traditional instrumental variables (IV) approach, a valid instrument has the following two properties: (1) it is strongly correlated with the language spoken at home, and (2) it is not correlated with test scores other than through language spoken at home.

One potential valid instrument is the county-level percentage of the population speaking a language other than English at home. Since this statistic represents the local language environment (such as local people's preference of speaking and keeping their origin languages), it is expected to be correlated with our treatment variable—the decision on whether or not to speak English at home. It should also not to be correlated with unmeasured determinants of test scores. However, this external instrument would be invalid if it is a proxy for unobserved family-level environment in that area, or a proxy for school quality in that county. For instance, in counties with a high percentage of Hispanic immigrants who usually do not speak English at home, schools may

[1] An instrumental variable approach can address the endogeneity caused by selection on unobservables and by reverse causality (see Sabia 2007).

provide second language (Spanish) instructions for students, which may affect children's cognitive test scores taken in English. Unfortunately, we cannot test the exogeneity of this instrument empirically when only one instrument is available. Given concerns about the validity of this external instrument and the difficulty in finding other good instruments, we adopt a new identification strategy proposed by Lewbel (2012).

### 3.2 *Lewbel IV Strategy*

Built on earlier work (Vella and Verbeek 1997; Cragg 1997; King et al. 1994; Klein and Vella 2003), Lewbel (2012) develops a new IV strategy that proposes restrictions on second moments and exploits heteroskedasticity in the errors for identification. The intuition for this technique is that the identification is achieved through internal instruments, which are constructed using the deviation from the mean of a vector of independent variables interacted with the residuals from a regression on the potential endogenous variable. The most important assumption of this IV strategy is that the error terms from that regression are heteroscedastic.

The Lewbel IV approach does not require an external instrumental variable and relaxes traditional assumptions by not requiring instruments to be unrelated to test scores. It has been widely used in cases in which other sources of identification such as external instruments are not available or are too weak (Sabia 2007; Kelly et al. 2014; and Millimet and Roy 2015). For example, Sabia (2007) adopts the Lewbel IV method to study the effect of body weight on adolescent academic performance given the concern that traditional instruments (biological mother's obesity status and biological father's status) might be correlated with the outcome (adolescent academic performance) directly. He first shows the existence of heteroskedasticy in the first-stage errors, and then constructs internal instruments using the deviation from the mean of all exogenous variables interacted with the residuals. His IV results are consistent with results from OLS models and individual fixed effects models.

Compared with conventional approaches, the Lewbel IV method exploits second moments to circumvent the need for instruments. It is useful when one is short of valid external instruments or traditional instruments are too weak (Mishra and Smyth 2015). The internal instruments can also be combined with external instruments to increase efficiency and be used to test for exogeneity (for example, see Mishra and Smyth 2015; Kelly et al. 2014). Therefore, in this paper, we first adopt a Lewbel IV approach, and then estimate an alternative model by supplementing these internal IVs with the external instrument to improve the strength of the instruments and maximize precision.

Lewbel (2012) IVs are constructed using the two following equations:

$$D_i = X_i\phi + \xi_i \tag{6}$$

$$Y_i = \alpha + \delta D_i + X_i\beta + u_i \tag{7}$$

where the language spoken at home ($D_i$) is determined by a set of observable characteristics, $X_i$, which also directly impact children's test scores (such child gender, race), and $Y_i$ is the observed test scores. In Equation (6), $\xi_i$ represents unobserved determinants of the language spoken at home, and it is correlated with $u_i$ in the presence of endogeneity.

The instruments are constructed as $[Z_i - E(Z_i)]\xi_i$, where $Z_i$ can be equal to $X_i$ or be a subset of $X_i$. In practice, since the components of $\xi_i$ cannot be observed, we replace them with residuals from the first-stage linear regression of $D_i$ on $X_i$. For a valid Lewbel IV approach, the following assumptions have to be satisfied: (1) $Z_i$ are uncorrelated with the covariance of $\xi_i$ and $u_i$ ($cov(Z_i, \xi_i, u_i) = 0$); (2) $Z_i$ are correlated with the variance of $\xi_i$ ($cov(Z_i, \xi_i^2) \neq 0$); and (3) $\xi_i$ is heteroskedastic. The first two assumptions are not testable, but they are standard assumptions and are unlikely to be major issues in practice. The third assumption is nonstandard, and Breusch-Pagan test is employed to examine the heteroskedasticity of errors terms.

Empirically, the heteroskedasticity could come from many sources, such as child race/ethnicity, school type, school region, and county of origin.[2] Here, we take child ethnicity as an example. As we know, on average, the pre-schools that Hispanic attend are not as good as those that non-Hispanic attend. Hispanic parents do not read to their children as often as non-Hispanic parents, and their children's nutrition are not as good as non-Hispanics'. As a results, Hispanic children may have a higher opportunity cost in learning than non-Hispanic children. The variance of errors may still the same within Hispanic children and non-Hispanic children, respectively, but differ considerably between these two groups.

The Lewbel IV model is estimated through the following process: (1) regress $D_i$ on $X_i$ and obtain error terms $e_i$, (2) construct internal instruments using $[Z_i - E(Z_i)]e_i$, (3) use these instruments to estimate the test score model via the two-stage least squares (2SLS) method.[3] It is important to note that the Lewbel estimates are potentially sensitive to the choice of $Z_i$. There are no standard rules for the optimal selection of $Z_i$. In this paper, following Millimet and Roy (2015), we select variables that are significantly correlated with the first-stage error variance (squares of residuals in practice).

In an alternative model, we supplement these internal instruments with our external instrument, county-level percentage of the population speaking a language other than English at home, to increase estimation efficiency. Internal instruments are constructed exactly the same way

[2] Baum (2006) says that in cross-sectional datasets, group membership is a common source of heteroskedasticity. Within each group, the assumption of homoskedasticity may still hold, but between groups, the variance usually differs. This group membership could be applied to workers in different industries or people from different states, for example.

[3] Note that equation (6) is not a first-stage model. It is a formular used to construct internal instruments.

as we did in the Lewbel IV approach. Equation (6) does not change. The only difference is that we use both internal IVs and external IV to estimate the test score model, which is Equation (7), via 2SLS. For both approaches, we have more than one exclusion restrictions for the potentially endogenous variable $D_i$. Therefore, overidentification tests are conducted empirically to examine whether it is appropriate to reject the null hypothesis that instruments are uncorrelated with the error terms of the test score estimation.

## 4. Data

Our data come from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, which is a nationally representative survey conducted by the U.S. Department of Education. It followed a cohort of approximately 20,000 students from kindergarten through middle school. Specifically, the students were interviewed in the fall and spring of kindergarten (1999), the fall and spring of first grade (1999), the spring of third grade (2002), the spring of fifth grade (2004), and the spring of eighth grade (2007). This survey has been widely used in the studies that investigate determinants of children's cognitive and non-cognitive development during early childhood (for example, Fryer and Levitt 2004; Claessens et al. 2009). It not only contains information about children's early schooling performance and social and emotional development, but also includes interviews with parents, teachers and schools, and thus provides a rich set of information on family background, educational resources at home, and school characteristics. The ECLS-K is particularly suited for this research because it asks parents whether the primary language spoken at home is English, and it has language screening tests that examine children's English proficiency. More importantly, the restricted use version of the data provides detailed information about the exact type of language spoken at home and the country of birth. It also contains the county identifiers, which enables us to merge this data with state- or county-level information to conduct the IV analysis.

In this paper, we focus on students in the third and fifth grades for two reasons. First, in the kindergarten year and first grade, a certain number of students from immigrant families (58 percent in the fall of kindergarten) did not pass the Oral Language Development Scale (OLDS) test, which is a language screening assessment administered to students with language minority background (speaking a language other than English at home). This screener is used to determine whether a child's English is good enough to receive cognitive tests in English. Only children who passed OLDS take math and reading tests.[4] To estimate the direct impact of language environment on test scores, we need to tease out the effect of English skills on academic performance and also keep as many language minority students as possible. Thus, we start from the third grade, when almost all children of immigrants demonstrated sufficient English proficiency to participate in the cognitive tests. Second, to empirically estimate the impact of language environment at home on early childhood development, we need at least two years of students' test scores, and the eighth grade is too late for an early childhood study.

### 4.1 *Key Dependent Variables*

Our primary outcome variables are reading and math standardized test scores. These two subjects reflect different dimensions of cognitive skills and thus allow us to test whether language spoken at home affects all types of cognitive skills equally. Specifically, reading tests measure

[4] Students who did not pass English OLDS but spoke Spanish would take a Spanish language screening test. For children who passed the Spanish version of OLDS, math and reading tests were administered in Spanish.

basic skills (such as print familiarity, letter recognition, and sounds), vocabulary and comprehension. Math tests examine student skills in counting and comparing numbers, solving problems and interpreting graphs, which should not require as high a level of English proficiency as reading. In this paper, we utilize item response theory (IRT) scale scores for each subject, which measure children's performance by incorporating the difficulty, discriminating ability, and guess-ability of each question. IRT scores have advantages over simple test scores which only count the number of correct answers (Fryer and Levitt 2004). Most importantly, IRT scoring allows for longitudinal measurement of gain and growth in achievement over time, which is another focus of this paper. We standardize the IRT scores to a mean of zero and a standard deviation of one for the sample on each of the tests and time periods.[5] This facilitates the comparison of our results to the impact of other important determinants on test scores examined in the literature.

### 4.2 *The Language Spoken at Home and Control Variables*

Our main independent variable is constructed from the survey question "what is the primary language spoken in your home," which is directly asked of parents.[6] We control for: (1) children's own characteristics, including race/ethnicity, gender, age (in months), and birth weight; (2) parental and household characteristics, including father's and mother's education, log family

[5] The raw IRT score is 0-212 for reading test, and 0-174 for math test. In our sample, reading test scores range from 51 to 201 with a mean of 130 and a standard deviation of 27 in the third grade, and 65 to 204 with a mean of 153 and a standard deviation of 25 in the fifth grade. Math test scores range from 35 to 166 with a mean of 102 and a standard deviation of 24 in the third grade, and 50 to 171 with a mean of 126 and a standard deviation of 24 in the fifth grade.

[6] This question was asked in children's kindergarten year and the first grade. We use the responses from the first grade because (1) the first grade responses have fewer missing values, and (2) some new students were added into the survey sample in the first grade, and their information was only available in the first grade. For those parents who replied to this question in both years, only 0.1% of them changed their answers, and we drop those observations.

income in children's kindergarten year,[7] [8] number of books at home, ownership of a computer at home, number of siblings, and parents' marital status; (3) school characteristics, including school type (public or private), school region (northeast, midwest, south, and west), school urbanicity (central city, urban fringe, or small town and rural areas), and a dummy variable indicating whether a student transferred from another school; and (4) child's country of birth (U.S. or another country) and time at entrance to the U.S. (before or after age 3).

The number of students participating in the ECLS-K survey is nearly 15,300 in the third grade and 11,800 in the fifth grade.[9] We drop observations without math and reading test scores, which is about 1 percent of the sample. Students who did not pass the OLDS test in the spring of first grade[10] and those with missing information on OLDS test are excluded as well (2.5 percent of the sample). We delete children who do not have information about the language spoken at home (4 percent of the sample). Observations with missing information on child, parent, and school characteristics are also dropped from the sample (17 percent of the sample).[11] The only exception is that observations with missing values on fathers' education are not deleted because a relatively large portion of the data (another 18 percent) has missing information on this variable. Instead, we

[7] Following Elder and Jepsen (2011), we use the log of family income to better capture the nonlinear effect of family income on test scores.

[8] In only the kindergarten year parents were asked about the exact income value. In the following years income was coded into $5,000 income brackets (top-coded at $200,000). To test whether the results are robust to the year and type of income controlled, we run specification checks later by controlling for income in the year of third and fifth grade (using a set of dummies). We find that the results are quite robust to different specifications of income.

[9] In this longitudinal survey, the sample of respondents decreases with each round of data collection. This is mainly because some children moved out of the original sample schools, and only a subsample of movers were followed.

[10] We realize that there might be a selection bias from dropping the students who did not pass the language screening test from the sample. We might underestimate a negative effect of speaking a language other than English at home (or overestimate a positive effect) by excluding students who had low English skills.

[11] Later we run specification checks for our baseline models by adding the observations with missing values on child, parent and school characteristics back into the sample. The results change little when we add back observations with missing values.

create a dummy variable indicating the missing status, equal to 1 if fathers' education is missing and zero otherwise. Children whose current geographical location cannot be identified or whose birth country is unknown are also deleted (1 percent of the sample). The final sample size is 10,521 for the third grade and 8,757 for the fifth grade.

### 4.3 *Summary Statistics*

Summary statistics of test scores are reported in Table 1. Simple comparisons of average test scores across different languages spoken at home reveal that students who speak English at home academically outperform those who do not in every aspect. The differences between these two groups are large and statistically significant in students' academic performance across both reading and math. In particular, the differences in reading test scores are as large as 0.44 and 0.45 standard deviations in the third and fifth grade, respectively, and the differences in math test scores are 0.29 and 0.22 in these two grades.[12] While the difference in reading scores increases slightly from the third to fifth grade, the difference in math scores decreases by a significant amount. Table 2 compares child, parent, and school characteristics between the group who speak English at home and the group who speak other languages at home using the sample of the third grade.[13] Examining the summary statistics of observable characteristics, we clearly see that students who speak English at home significantly differ from those who do not in most dimensions. This table highlights the importance of controlling for observables in our estimations. More importantly, it highlights the potential endogeneity problem, discussed above, that students speaking English at home may differ

[12] In terms of raw scores, the differences in reading test scores are 11.88 and 11.25 in the third and fifth grade, which are about 9 percent and 7 percent of their means. The differences in math test scores are 6.96 and 5.28 in the third and fifth grade, which are about 7 percent and 4 percent of their means.

[13] Appendix Table (1) presents the summary statistics of the control variables for the sample of the fifth grade.

from students speaking other languages at home in unobservable dimensions. This issue may prevent us from drawing a causal interpretation from the OLS results.

### 4.4 *External Instrumental Variable*

The external instrumental variable used is the county-level percentage of the population (5 years and older) speaking a language other than English at home. This information is collected in the American Community Survey (ACS), which is conducted annually by Census Bureau. The ACS asks each respondent the primary language spoken at home and provides county-level rates of a language other than English being spoken at home. We merge this data with our sample by using the confidential county-level geographic information in ECLS-K. Since the ACS only provides summary estimates for counties with 250,000 or more people in a population, we lose a part of the sample by merging these two data sets. As a result, only children from large counties are included in the baseline IV estimation. Later we use a similar state-level instrument to test the robustness of our results.

## 5. Results

### 5.1 *OLS Results on the Level of Test Scores*

We begin with a parsimonious set of covariates, and then expand that set of covariates. The results are reported in Table 3. Panel A presents estimates on test scores in the third grade. As shown in Column 1, for the model including only children's basic demographic information and birth weight, we find that a non-English-language environment at home is correlated with a large, significant decrease in both reading and math scores. This is consistent with the raw differences of test scores between the group speaking English at home and the group not speaking English presented in Table 1. We then include family and school characteristics sequentially into the specification. Information on country of birth and time of child immigration is captured in the last specification. As we can see in Columns 2, 3 and 4, the negative association, albeit smaller, remains sizable and significant for reading tests. In particular, in the model with a full set of controls, we find that speaking a language other than English at home is associated with a decrease of 0.34 standard deviation in reading scores. By contrast, the association between a non-English-language environment at home and math scores becomes much smaller and statistically insignificant once we control for family characteristics.

Turning to the estimates on test scores in Grade 5, reported in Panel B of Table 3, we again see a similar pattern. Across all specifications, speaking a language other than English at home is significantly correlated with a decrease in reading scores. By comparing the estimates in the third grade to those in the fifth grade, we find that this negative correlation for reading test does not shrink or disappear over time. For the math test, we find a negative association between a non-English-language home environment and test scores in the model controlling for only child characteristics. The significance of this negative association disappears while we control more variables.

That the effect varies across specifications implies that these additional control variables are highly correlated with the type of languages spoken at home, which is consistent with the literature. By summarizing prior studies (Crawford 1997; Moss and Puma 1995), Hernandez (2007) finds that language-minority students are most likely to attend schools that have a high percentage of poor students and inexperienced teachers, and that they are over-represented in special education classes. Thus, it is important to control for these family and school characteristics so as to identify a causal effect. However, even after conditioning on such a rich set of covariates, we may still fail to control for individual heterogeneity adequately. To address this endogeneity issue, we turn to IV estimates.

### 5.2 *IV Results on the Level of Test Scores*

Given that we cannot test the exogeneity, and therefore, the validity of our external instrument, and also to avoid the potential weak external instrument issue, we employ an internal IV approach proposed in Lewbel (2012). One important assumption of this approach is the presence of heteroskedastic errors in estimation of Equation (6). To test this assumption, we conduct the Breusch-Pagan test. As shown in the first two columns of Table 4, we are able to reject the null hypothesis of homoskedastic errors at the $p \leq 0.0001$ level for models in both the third and the fifth grades.[14] To construct internal instruments, we select variables that are significantly related to the variance of the errors in Equation (6). For the third grade sample, race/ethnicity (Hispanic, Asian, other), family income, school region (Midwest, South, West), school urbanicity (small town), and age at the entrance to the U.S. (older than three) are included in $Z_i$; and for the fifth grade sample, race/ethnicity (Hispanic, Asian, other), family income, ownership of computer, school region (South), school urbanicity (small town), and age at the entrance to the U.S. (older than three) are chosen.

Before proceeding to the IV estimation, we test whether these internal instruments are strongly correlated with the choice of language spoken at home and whether they are exogenous to test scores. Results are presented in Columns 1 and 2 of Table 4. A simple F test on excluded

[14] In each grade, Equation (6) is the same for reading and math scores. Therefore, the heteroskedasticity result is the same for both subjects, and we report only once for each grade.

instruments is first performed to test for weak instruments. A rule of thumb of this test is that the F should be above 10 in order to reject the null of weak instruments (Staiger and Stock 1997). As shown in the first two columns, our F statistics are much larger than 10, indicating strong instruments. Another more formal weak identification test—the Gragg Donald test—is conducted as well, and the Kleibergen-Paap Wald rk F statistic is presented. According to Stock-Yogo critical values, we are able to reject the null hypothesis that instruments are weakly associated with the language spoken at home at $p \leq 0.0001$ level. Additionally, we perform an overidentification test to examine whether the excluded instruments are correlated with our outcomes: reading and math test scores. The Hansen J statistic and p-value shown in the Table 4 indicate a failure to reject the null hypothesis that the exclusion restriction is invalid. In other words, the exogeneity assumption is satisfied; our instruments do not appear to be correlated with test scores. Based on these test results, we argue that the internal instruments we construct have the important properties of good instruments.

Panels A and B in Column 1 of Table 5 presents Lewbel IV estimates of the effect of a non-English-language home environment on test scores.[15] As shown in Panel A, speaking a language other than English at home reduces reading scores by 0.14 standard deviation[16], while has no significant effect on math scores in the third grade. In the fifth grade, again, we can see that speaking a language other than English at home is associated with a decrease in reading scores and the magnitude is 0.17 standard deviation (shown in Panel B)[17]. The estimate for math scores is not

[15] Appendix Table (2) presents first-stage estimation results.

[16] In terms of raw scores, this is a reduction of 3.78 at the mean of 130, which is about 3 percent of its mean.

[17] In terms of raw scores, this is a reduction of 5.28 at the mean of 126, which is about 3 percent of its mean.

statistically significant. These results are consistent with OLS results, but are relatively larger, which is common for IV estimates.

To improve statistical power, we further supplement the internal instruments with an external instrument: the county-level percentage of the population speaking a language other than English at home. Columns 3 and 4 of Table 4 report the coefficient on the external IV in the first-stage model, and present results of various diagnostic tests on the validity of instruments. As shown, the statistically significant association between this county-level instrument and our potential endogenous regressor in both third and fifth grades indicates that local language environment is strongly related to an individual's choice of language spoken at home. Results of Breusch-Pagan tests provide us with evidence that the errors are heteroskedastic. Based on the F statistics shown in the table, we do not find the evidence of weak instruments. The overidentification test statistics indicate that our instruments are correlated with neither reading nor math scores for both grades.

The second-stage results for this internal-plus-external IV estimation are presented in Panel A and Panel B of Table 5, Column 2.[18] We continue to find a negative effect of a non-English-language environment at home on reading test scores across both third and fifth grades, but no significant effect on math scores in either grades. In the third grade, speaking a language other than English at home decreases student reading scores by 0.15 standard deviation, and in the fifth grade, this negative effect is about 0.16 standard deviation.

Overall, both Lewbel IV approach and county IV-plus-Lewbel IV strategy provide us valid instruments and allow us to identify a causal effect of language environment at home on child cognitive development. The estimates from both models are qualitatively similar to those from an OLS model.

---

[18] Appendix Table 3 presents full first-stage results when using county IV-plus-Lewbel IV to identify the casual effect.

### 5.3 *Results on the Growth of Test Scores*

In earlier models, we only focus on the effect of speaking a language other than English at home on levels of test scores. This tells us that speaking non-English languages at home decreases reading test scores, but provides no information about whether the negative effect increases or decreases over time (or whether it leads to a larger or smaller test score gap over time). In addition to the effect on levels of test scores, we are also interested in knowing whether a non-English-language environment at home has any effect on the growth of test scores. If the longer time a child is exposed to a bilingual environment, the larger this detrimental effect is, we expect to see a negative association between growth of test scores and speaking non-English languages at home. However, if the negative effect decreases when a child is more fluent in both languages, we would see a positive effect on growth of test scores.

To test which argument is supported by our data, we estimate the effects of languages spoken at home on: (1) the first difference in test scores between the third and fifth grades, and (2) the growth rate of test scores from the third to fifth grade. Columns 4-8 of Table 4 report diagnostic test statistics on instruments for each model. The results illustrate that our instruments are valid. Panel C and Panel D in Table 5 present IV estimation results. Both Lewbel IV results and county IV-plus-Lewbel IV results fail to provide any evidence that language environment at home influences growth of test scores. Therefore, the negative effect on reading scores in the fifth grade appears to be a result of the gap in the third grade. In other words, although language environment at home has a long-lasting effect on reading scores in early childhood, the effect does not increase (or the cumulative effect does not grow) over time.

Our result suggests that language skills might not be the main mechanism through which language environment at home affects test scores. If English skill is the main mechanism, with an improvement in English among language minority students from the third to fifth grade, we would expect a decrease in the test score gap, or a positive effect on growth of test scores.

### 5.4 *Specification Checks*

We conduct several specification checks to examine the robustness of our results. By merging our county-level instrument collected by the ACS with ECLS-K data, we lose children living in small counties (pop<250,000). Therefore, our county IV-plus-Lewbel IV model is not estimated on a full sample. In order to check whether excluding those children would affect our results, we use the state-level percentage of the population speaking a language other than English at home augmented with Lewbel internal IVs to estimate all the models again. The results on levels of test scores and growth of scores are presented in the first column of Table 6.[19] As shown, speaking a language other than English at home has a significant negative effect on reading scores in both third and fifth grades. The magnitude of the negative effect is similar to that found in county IV-plus-Lewbel IV estimation. For math scores, the effect is still insignificant in both grades. We consistently find no significant effect on growth of test scores. We further restrict the sample by including only observations in large counties (pop>250,000) and estimate OLS models, Lewbel IV models, and state IV-plus-Lewbel IV models again. We report results for state IV-plus-Lewbel IV estimation in Column 2 of Table 6. For both levels of scores and growth of scores, results are

[19] In the first-stage estimation, state-level language environment is also strongly associated with an individual's use of language at home. Results on the first-stage estimation and diagnostic tests are not reported here, but are available upon request.

largely unchanged. The OLS results and Lewbel IV results are also similarly to the results using a full sample.

Next, we exclude students who were not born in the U.S. (about 300 observations in the full sample) to see whether our results are affected. Compared with U.S. born children, those who came to U.S. at an early age are more likely to experience a culture shock and face obstacles in adjusting to a new environment, especially those from non-English speaking countries. Therefore, we want to see whether our results are sensitive to this group of people. Column 3 of Table 6 reports the county IV-plus-Lewbel IV results using a sample of the U.S. born children. As shown, the results are not sensitive to whether or not we include non-U.S. born children.

Overall, we find a robust, consistent negative effect of speaking non-English languages at home on reading scores, but such a detrimental effect does not exist for math scores.

## 6. Subsample Analysis and Results

### 6.1 *Spanish versus Other Languages*

Since Hispanics are the largest racial/ethnic minority in the U.S. (accounted for 47 percent of the foreign-born population and 14.5 percent of total population), Spanish is the most commonly spoken language other than English in the U.S. According to the American Community Survey in 2013, among people who use a language other than English at home, more than 60 percent speak Spanish. We are interested to know the effect of a non-English-language language environment at home for this particular group not only because a majority of immigrant children speak Spanish, but also because the effect may differ between children speaking Spanish and children speaking other foreign languages. Children speaking Spanish at home are more likely to be placed in a bilingual education program in schools than children speaking other languages (Zehler et al. 2003, Aimee Chin 2015). By far the most common form of bilingual education is Spanish-English. Spanish-speaking limited English proficient students are the most numerous, and they are less geographically dispersed than other immigrants. These factors could make a non-English language environment at home play a different role for Spanish-speaking children and other foreign-language-speaking children.

We first restrict the sample to children whose primary language spoken at home is either Spanish or English so as to explore the effect of speaking Spanish at home on test scores. Columns 1 and 2 of Table 7 report the results of Lewbel estimation and county IV-plus-Lewbel IV estimation.[20] We find that speaking Spanish rather than English at home significantly reduces reading scores while it has no effect on math scores in both third and fifth grades. This is consistent with the full sample analysis.

Next, we estimate the effect of a non-English language at home on test scores for children who speak non-Spanish at home. We drop children whose primary language at home is Spanish from the sample and re-estimate all models again. The results are presented in Columns 3 and 4 of Table 7. Similar to the results we find in full sample analysis, speaking a language other than English at home decreases reading scores, but has no significant effect on math scores. The results shown in the first four columns of Table 7 indicate that the effect of speaking a language other than English at home on test score differs little by the type of foreign language spoken at home.[21]

### 6.2 *Boys versus Girls*

[20] Appendix Table (4) reports the first-stage IV estimation results and diagnostic test statistics on the validity of instruments.

[21] Spanish is the only language that we have enough observations to separate it from other non-English languages. The sample size for other foreign language users are very small, and we are not able to test the effect by each non-English language.

To test whether the effect of language environment at home on child test scores differs by gender, we stratify the sample by gender. Results are presented in Columns 5-8 of Table 7.[22] In both samples of boys and girls, our results show that speaking a language other than English at home has a negative effect on reading scores in the third and fifth grades. For math scores in boys, we find that the negative association between a non-English-language home environment and test scores in the third grade turns into a positive association in the fifth grade, although neither estimate is statistically significant. For math scores in girls, we find that the effect of a non-English-language environment at home on test scores is negative in both grades. As shown in Column 7, speaking a language other than English at home decreases math scores by 0.15 standard deviation in the third grade, and 0.14 standard deviation in the fifth grade. Overall, the negative effect of a non-English-language environment at home on reading test scores differs little by gender, but its effect on math scores does vary between girls and boys. It seems that boys' math scores are not affected by the language used at home, while girls' math scores are negatively affected by a non-English-language environment at home.

## 7. Conclusion

In this paper, we examine the effect of speaking a language other than English at home on students' academic achievement during early childhood. By controlling for a rich set of child, family and school characteristics, our OLS results indicate that the use of a non-English language at home is negatively associated with reading test scores in the third grade and this negative

[22] Appendix Table (5) reports the first-stage IV estimation results and diagnostic test statistics on the validity of instruments.

association persists to the fifth grade. However, there is no significant effect on math scores. These results remain qualitatively unchanged even after we correct for the potential endogeneity. Both our Lewbel IV results and external-plus-internal IV results indicate that speaking a language other than English at home decreases reading scores, and the magnitude ranges from 0.14 to 0.17 standard deviation. The external IV we used is county-level rates of speaking a language other than English at home. Further, we do not find any evidence that the use of a language other than English at home affects the growth rate of test scores. This indicates that the negative effect on reading test scores appears to be a result of the initial gap in the third grade, which is constant over time. All the results are robust to a similar set of instruments measured at the state-level.

By stratifying the sample into several subgroups, we find that the pattern in the effects of language environment remains unchanged whether the foreign language is Spanish or not. We also find that the effect differs by gender. The language environment appears to have a negative effect on both reading and math scores among girls, while it only has a negative effect on reading scores for boys. Overall, our results provide the evidence that early language environment has an important effect on the cognitive development of 3$^{rd}$ graders, and this effect persists through their childhood.

In sum, these results seem more line with the direct effect of bilingualism on academic performance suggested by the subtractive view in the psychology literature. The negative effect on reading scores in the third and fifth grades is also consistent with lack of language proficiency for English learners. However, since we do not find any significant effect on growth of test scores, the language-skill mechanism seems not to be the main mechanism unless there is no improvement on English for language minority students from the third to fifth grade.[23] The fact that all of students in our sample have passed the English screening test before Grade 3 provides another argument that English skill is not the main mechanism.

Closing the achievement gap between children of immigrants and their native peers has been a long-standing goal for educators. Up to date, most policy makers and researchers have been focusing on evaluating the effectiveness of various education programs in schools, while relatively less attention has been put on family environment. Our study provides evidence that the language environment at home has a significant effect on academic achievement during early childhood and suggests for potential policy interventions.

The issue of how to help language minority students is also at the center stage of education policy debates. In earlier years between 1970 and 1990, both federal and some individual states had laws mandating bilingual education programs in certain schools to address the educational needs of limited English proficiency students. However, since the late 1990s, there has been a shift away from using bilingual education toward using English-only programs to help these students. In the last 15 years, two thirds states have eliminated bilingual education programs in public schools. Our results contribute to this debate by providing some indirect evidence that learning and speaking two or more languages has a negative effect on academic performance at early childhood, and support for using English-only programs to help language minority students.

[23] This could be true if language minority students are more concentrated in certain schools, and mainly talk and play with students speaking the same foreign language.

## LIST OF REFERENCES


Abutalebi, Jubin and Harald Clahsen. 2015. “Bilingualism, Cognition, and Aging.” *Bilingualism: Language and Cognition*, 18: 1-2.

Baum, Christopher F. 2004 “Stata Tip 38: Testing for Groupwise Heteroskedasticity.” *The Stata Journal*, 6 (4): 590-592.

August, D. and K. Hakuta (eds.). 1997. *Improving Schooling for Language-Minority Children: A Research Agenda*. Washington: D.C.: National Academy Press.

Bialystok, Ellen. 1999. “Cognitive Complexity and Attentional Control in the Bilingual Mind.” *Child Development*, 70: 636-644.

Bialystok, Ellen. 2007. “Cognitive Effects of Bilingualism: How Linguistic Experience Leads to Cognitive Change.” *International Journal of Bilingual Education and Bilingualism*, 10: 210-223.

Bialystok, Ellen. 2009. “Bilingualism: The good, the Bad, and the Indifferent.” *Bilingualism: Language and Cognition*, 12: 3-11.

Bialystok, E. 2013. “25 The Impact of Bilingualism on Language and Literacy Development.” *In The Handbook of Bilingualism and Multilingualism*.

Bialystok, Ellen, Fergus IM Craik, and Gigi Luk. 2012. “Bilingualism: Consequences for Mind and Brain.” *Trends in Cognitive Sciences*, 16: 240-250.

Bialystok, E. and M. M. Martin. 2004. “Attention and Inhibition in Bilingual Children: Evidence from the Dimensional Change Card Sort Task.” *Developmental Science*, 7: 325-339.

Bleakley, H. and A. Chin. 2004. “Language Skills and Earnings: Evidence from Childhood Immigrants.” *Review of Economics and Statistics*, 86: 481-496.

Cameron, S. and J. Heckman. 2001. “The Dynamics of Educational Attainment for Black, Hispanic, and White Males.” *Journal of Political Economy*, 109: 455-499.

Chin, Aimee. 2015. “Impact of Bilingual Education on Student Achievement.” *IZA World of Labor,* 131.

Chin, Aimee, N. Meltem Daysal, and Scott A. Imberman. 2013. “Impact of Bilingual Education Programs on Limited English Proficient Students and Their Peers: Regression Discontinuity Evidence from Texas.” *Journal of Public Economics*, 107:63-78.

Chiswick, B.R. and P.W. Miller. 2008. “A Test of the Critical Period Hypothesis for Language Learning.” *Journal of Multilingual and Multicultural Development*, 29: 16-29.

Chiswick, R., Y. Lee, and P. Miller. 2005. “Parents and Children Talk: English Language Proficiency within Immigrant Families.” *Review of the Economics of Household*, 3: 243-268.

Claesens, A., G. Duncan, and M. Engel. 2009. “Kindergarten Skills and Fifth-Grade Achievement: Evidence from the ECLS-K.” *Economics of Educational Review*, 28: 415-427.

Colding, B., L. Husted, and H. Hummelgaard. 2009. "Educational Progression of Second generation Immigrants and Immigrant children." *Economics of Education Review*, 28: 434-443.

Cook, V. 1997. *Tutorials in Bilingualism: Psycholinguistic Perspective*, chapter The Consequences of Bilingualism for Cognitive Processing. Mahwah, NJ: Lawrence Erlbaum Associates.

Cragg, John G. 1997. "Using Higher Moments to Estimate the Simple Errors-in-Variables Model." The *RAND Journal of Economics*, pp. S71-S91.

Crawford, James. 1997. "Best Evidence: Research Foundations of the Bilingual Education Act." Technical report, Washington, DC: National Clearinghouse for Bilingual Education.

Diaz, R.M. 1983. "Thought and Two Languages: The Impact of Bilingualism on Cognitive Development." *Review of Research in Education*, 10: 23-54.

Dustmann, C. and A. Glitz. 2011. *Handbook of the Economics of Education*, volume 4, chapter Migration and Education.

Dustmann, C., S. Machin, and U. Schonberg. 2010. "Ethnicity and Educational Achievement in Compulsory Schooling." *The Economic Journal*, 120: 272-297.

Dustmann, Christian and Claudia Trentini. 2008. "Ethnic Test Score Gaps: Integration and Language Skills of Kindergarten Children in England." Unpublished Manuscript. http://www.homepages.ucl.ac.uk/~uctpctr/dustmann_trentini.pdf

Elder, T. and C. Jepsen. 2011. "Are Catholic Primary Schools More Effective Than Public Primary Schools." Unpublished Manuscript. https://www.msu.edu/~telder/Cath_Prim_Current.pdf

Fortuny, Karina and Ajay Chaudry. 2009. "Children of Immigrants—Immigration Trend." Technical report, Urban Institute.

Fryer, R.G. and S.D. Levitt. 2004. "Understanding the Black-White Test Score Gap in the First Two Years of School." *The Review of Economics and Statistics*, 86:447-464.

Greene, Jay P. 1998. "A Meta-Analysis of the Effectiveness of Bilingual Education." Technical report, Tomas Rivera Policy Institute.

Greenspan, S. 1997. *The Growth of the Mind. Reading*, MA: Addison-Wesley.

Hakuta, K. 1986. "Cognitive Development of Bilingual Children." Non-journal, California Univ., Los Angeles. Center for Language Education and Research.

Heckman, J. 2008. "Schools, Skills, and Synapses." *Economic Inquiry*, 46: 289-324.

Hernandez, C. 2007. "Home Language Use and Hispanic Academic Achievement: Evidence from Texas High Schools." *Penn McNair Research Journal* 1: Article 4.

Homel, Peter, Michael Palij, and Doris Aaronson (eds.). 2014. Childhood Bilingualism: *Aspects of Linguistic, Cognitive, and Social Development*. Psychology Press.

Keane, M. and K. Wolpin. 1997. "The Career Decisions of Young Men." *Journal of Political Economy*, 105: 473-522.

Keane, M. and K. Wolpin. 2001. "The Effect of Parental Transfers and Borrowing Constrains on Educational Attainment." *International Economic Review*, 42:1051-1103.

Kelly, Inas Rashad, Dhaval M. Dave, Jody L. Sindelar, and William T. Gallo. 2014. "The Impact of Early Occupational Choice on Health Behaviors." *Review of Economics of the Household*, 12:737-770.

King, M., E. Sentana, and S. Wadhwani. 1994. "Volatility and Links between National Stock Markets." *Econometrica*, 62: 901-923.

Klein, Roger and Francis Vella. 2003. "Identification and Estimation of the Triangular Simultaneous Equations Model in the Absence of Exclusion Restrictions through the Presence of Heteroskedasticity." Unpublished Manuscript. http://www.econ.ku.dk/CAM/Files/Spring 2004/revnotes.pdf

Lewbel, A. 2012. "Using Heteroskedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models." *Journal of Business & Economic Statistics*, 30(1).

Matsudaira, Jordan D. 2005. "Sinking or Swimming? Evaluating the Impact of English Immersion versus Bilingual Education on Student Achievement." Technical report, University of California, Berkeley.

Millimet, Daniel L. and Jayjit Roy. 2015. "Empirical Tests of the Pollution Haven Hypothesis When Environmental Regulation is Endogenous." *Journal of Applied Econometrics*.

Mishra, Vinod and Russell Smyth. 2015. "Estimating Returns to Schooling in Urban China Using Conventional and Heteroskedasticity-Based Instruments." *Journal of Applied Econometrics*, 47:166-173.

Moss, M. and M. Puma. 1995. *Prospects: The Congressionally Mandated Study of Educational Growth and Opportunity.* First Year Report on Language Minority and Limited English Proficient Students. 1118 22nd Street, N.W., Washington, DC 20037: The National Clearinghouse for Bilingual Education at George Washington University.

Pelham, Sabra D. and Lise Abrams. 2014. "Cognitive Advantages and Disadvantages in Early and Late Bilinguals." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40: 313.

Pope, D.G. 2008. "Benefits of Bilingualism: Evidence from Mormon Missionaries." *Economics of Education Review*, 27: 234-242.

Rampney, B.D., G.S. Dion, and P.L. Donahue. 2009. "NAEP 2008 Trends in Academic Progress." Technical Report NCES-2009-479, National Center for Education Statistics, Institute of Education Sciences. U.S. Department of Education, Washington, DC.

Sabia, J. Joseph. 2007. "The Effect of Body Weight on Adolescent Academic Performance." *Southern Economic Journal*, 4:871-900.

Schnepf, S.V. 2007. "Immigrants' Educational Disadvantage: An Examination across Ten Countries and Three Surveys." *Journal of Population Economics*, 20:527-545.

Slavin, Robert E., Nancy Madden, Margarita Calderon, Anne Chamberlain, and Megan Hennessy. 2011. "Reading and Language Outcomes of a Five-Year Randomized Evaluation of Transitional Bilingual Education." *Educational Evaluation and Policy Analysis*, 33:47-58.

Staiger, Douglas and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica*, 65:557-586.

Stock, J. and M. Yogo. 2005. *Testing for Weak Instruments in Linear IV Regression*, pp. 80-108. New York: Cambridge University Press.

Vella, F. and M. Verbeek. 1997. "Order as an Instrumental Variable: An Application to the Return to Schooling." Unpublished paper.

Yoshida, Hanako. 2008. "The Cognitive Consequences of Early Bilingualism." Zero to Three 11: 26-30.

Zehler, A. M., H. L. Fleischman, P. J. Hopstock, T. G. Stephenson, M. L. Pendzick, and S. Sapru. 2003. "Descriptive Study of Services to LEP Students and LEP Students with Disabilities." Volume IA Research Report — Text. Washington, DC: US Department of Education.

**Table 1. Summary Statistics (Test Scores)**

| Variables | Language Environment at Home | | Difference |
|---|---|---|---|
| | English | Non-English | |
| | (1) | (2) | (3)=(1)-(2) |
| **Panel A: Grade 3 Test Scores** | | | |
| Reading | 0.05 | -0.39 | 0.44*** |
| | (0.99) | (0.97) | |
| Mathematics | 0.03 | -0.26 | 0.29*** |
| | (1.00) | (0.98) | |
| Science | 0.07 | -0.54 | 0.61*** |
| | (0.99) | (0.91) | |
| N | 9371 | 1150 | |
| **Panel B: Grade 5 Test Scores** | | | |
| Reading | 0.06 | -0.39 | 0.45*** |
| | (0.99) | (1.01) | |
| Mathematics | 0.03 | -0.19 | 0.22*** |
| | (0.99) | (1.04) | |
| Science | 0.07 | -0.45 | 0.52*** |
| | (0.98) | (1.04) | |
| N | 7671 | 1086 | |

Note: 1. Standard deviations are in parentheses. ***, **, and * indicate the estimates are statistically significant at the 1%, 5%, and 10% levels, respectively. All test scores are standardized to a mean of zero and a standard deviation of one.
Source: Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K).

## Table 2. Summary Statistics (Control Variables)

| Variables | Control Set | Language Environment at Home<br>English<br>(1) | <br>Non-English<br>(2) | Difference<br>(3)=(1)-(2) |
|---|---|---|---|---|
| | | **Grade 3 Sample** | | |
| **Race/Ethnicity** | | | | |
| White | A | 0.71 | 0.07 | 0.64*** |
| | | (0.46) | (0.25) | |
| Black | A | 0.12 | 0.01 | 0.11*** |
| | | (0.32) | (0.09) | |
| Hispanic | A | 0.10 | 0.62 | -0.52*** |
| | | (0.29) | (0.49) | |
| Asian | A | 0.03 | 0.27 | -0.25*** |
| | | (0.16) | (0.44) | |
| Other | A | 0.06 | 0.03 | 0.02*** |
| | | (0.23) | (0.18) | |
| **Male** | A | 0.50 | 0.52 | -0.02 |
| | | (0.50) | (0.50) | |
| **Age (in months)** | | | | |
| less than 105 months | A | 0.07 | 0.10 | -0.04*** |
| | | (0.25) | (0.31) | |
| 105-107 months | A | 0.20 | 0.22 | -0.01 |
| | | (0.40) | (0.41) | |
| 108-110 months | A | 0.23 | 0.22 | 0.01 |
| | | (0.42) | (0.41) | |
| 111-113 months | A | 0.23 | 0.27 | -0.04*** |
| | | (0.42) | (0.44) | |
| 114-116 months | A | 0.18 | 0.15 | 0.01*** |
| | | (0.39) | (0.36) | |
| 117 or more months | A | 0.09 | 0.04 | 0.05*** |
| | | (0.28) | (0.20) | |
| **Birth Weight (1 b)** | A | 6.99 | 6.86 | 0.13*** |
| | | (1.32) | (1.27) | |
| **Father's Education** | | | | |
| Less than High School | B | 0.05 | 0.29 | -0.24*** |
| | | (0.23) | (0.45) | |
| High School | B | 0.21 | 0.2 | 0.01 |
| | | (0.41) | (0.40) | |
| Some College | B | 0.24 | 0.17 | 0.07*** |
| | | (0.43) | (0.37) | |
| Bachelor and Above | B | 0.31 | 0.21 | 0.10*** |
| | | (0.46) | (0.41) | |
| Education Missing | B | 0.19 | 0.13 | 0.05*** |
| | | (0.39) | (0.34) | |
| **Mother's Education** | | | | |
| Less than High School | B | 0.06 | 0.33 | -0.27*** |
| | | (0.23) | (0.47) | |
| High School | B | 0.24 | 0.23 | 0.01 |
| | | (0.43) | (0.42) | |

## Cont. Table 2. Summary Statistics (Control Variables)

| Variables | Set | Language Environment at Home<br>English<br>(1) | <br>Non-English<br>(2) | Difference<br>(3)=(1)-(2) |
|---|---|---|---|---|
| | | **Grade 3 Sample** | | |
| **Mother's Education** | | | | |
| Some College | B | 0.38 | 0.23 | 0.14*** |
| | | (0.49) | (0.42) | |
| Bachelor and Above | B | 0.32 | 0.2 | 0.12*** |
| | | (0.47) | (0.40) | |
| **Log Family Income (in 1998)** | B | 10.63 | 9.97 | 0.66*** |
| | | (1.50) | (1.94) | |
| **Numbers of Books at Home** | B | 137.34 | 61.66 | 75.68*** |
| | | (189.35) | (102.96) | |
| **Own a computer at Home** | B | 0.85 | 0.67 | 0.18*** |
| | | (0.36) | (0.47) | |
| **Number of Siblings** | B | 1.52 | 1.87 | -0.35*** |
| | | (1.06) | (1.34) | |
| **Single Parents** | B | 0.19 | 0.13 | 0.06*** |
| | | (0.39) | (0.34) | |
| **Public School** | C | 0.78 | 0.91 | -0.13*** |
| | | (0.42) | (0.29) | |
| **Transfer Student** | C | 0.16 | 0.19 | -0.03*** |
| | | (0.37) | (0.39) | |
| **School Region** | | | | |
| Northeast | C | 0.20 | 0.17 | 0.03*** |
| | | (0.40) | (0.38) | |
| Midwest | C | 0.29 | 0.13 | 0.16*** |
| | | (0.45) | (0.33) | |
| South | C | 0.32 | 0.24 | 0.09*** |
| | | (0.47) | (0.43) | |
| West | C | 0.19 | 0.46 | -0.28*** |
| | | (0.39) | (0.50) | |
| **School Urbanicity** | | | | |
| Central City | C | 0.34 | 0.58 | -0.24*** |
| | | (0.47) | (0.49) | |
| Urban and Large Town | C | 0.40 | 0.35 | 0.06*** |
| | | (0.49) | (0.48) | |
| Small Town and Rural | C | 0.25 | 0.07 | 0.18*** |
| | | (0.44) | (0.26) | |
| **Born in the U.S.** | D | 0.99 | 0.89 | 0.01*** |
| | | (0.10) | (0.31) | |
| **Came to the U.S.** | D | 0.00 | 0.06 | 0.05*** |
| **After Age 3** | | (0.05) | (0.23) | |
| **N** | | 9599 | 1213 | |

Note: Standard deviations are in parentheses. ***, **, and * indicate the estimates are statistically significant at the 1%, 5%, and 10% levels, respectively. Control set A includes all variables listed in set A. Control set B includes variables in control set A and variables listed in set B. Control Set C includes variables in control set B and variables listed in set C. Control set D includes variables in control set C and variables listed in set D.
Source: Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K).

**Table 3. OLS Results: the Effect of Speaking A Language Other Than English at Home on Test Scores**

| | OLS Estimates | | | |
|---|---|---|---|---|
| | **(1)** | **(2)** | **(3)** | **(4)** |
| **Panel A: Grade 3 Test Scores** | | | | |
| Reading Tests | -0.342*** | -0.080** | -0.081** | -0.081** |
| | (0.070) | (0.032) | (0.032) | (0.033) |
| Math Tests | -0.218** | 0.005 | -0.013 | -0.016 |
| | (0.075) | (0.045) | (0.046) | (0.045) |
| N | 10521 | 10521 | 10521 | 10521 |
| **Panel B: Grade 5 Test Scores** | | | | |
| Reading Tests | -0.358*** | -0.108** | -0.114*** | -0.124*** |
| | (0.074) | (0.040) | (0.036) | (0.036) |
| Math Tests | -0.164** | 0.049 | 0.027 | 0.019 |
| | (0.066) | (0.036) | (0.035) | (0.035) |
| N | 8757 | 8757 | 8757 | 8757 |
| **Control Sets** | A | B | C | D |

Note: 1. Robust standard errors are in parentheses. ***, **, and * indicate the estimates are statistically significant at the 1%, 5%, and 10% levels, respectively.
2. Control set A includes children's characteristics: race/ethnicity, gender, age (in months), birth weight; control set B includes the variables in control set A and household characteristics: father's education, mother's education, log family income in the kindergarten year, number of books at home, ownership of a computer, number of siblings, marital status; control set C includes the variables in control set B and school characteristics: attending public school, indicator of a transfer student, school region, and school urbanicity; control set D includes the variables in control set C and born in the U.S. and entered into the U.S. after age 3.
Source: Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K).

## Table 4. Diagnostic Tests for Instruments in IV Estimation

(Dependent Variable: Speaking a Language Other Than English at Home)

| | Lewbel IV (1) | Lewbel IV (2) | County IV +Lewbel IV (3) | County IV + Lewbel IV (4) | Lewbel IV (5) | Lewbel IV (6) | County IV +Lewbel IV (7) | County IV + Lewbel IV (8) |
|---|---|---|---|---|---|---|---|---|
| | **Estimation on Level of Test Scores** | | | | **Estimation on Growth of Test Scores** | | | |
| | **Grade 3** | **Grade 5** | **Grade 3** | **Grade 5** | **First Difference** | **Growth Rate** | **First Difference** | **Growth Rate** |
| **Coefficients on External Instruments** | | | | | | | | |
| % Population Speaking Non-English at Home | | | 0.116*** | 0.132*** | | | 0.114*** | 0.114*** |
| | | | (0.027) | (0.031) | | | (0.027) | (0.027) |
| **Heteroskedasticity Test** | | | | | | | | |
| Breusch-Pagan Test Stat. | 9528.316 | 6925.157 | 3298.517 | 2250.292 | 6546.391 | 6546.391 | 2148.215 | 2148.215 |
| Breusch-Pagan p Value | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **Weak Instruments Tests** | | | | | | | | |
| F Stat. on Instruments | 613.444 | 595.876 | 596.817 | 416.336 | 1674.034 | 1674.034 | 495.897 | 495.897 |
| Kleibergen-Paap F stat. | 1154.114 | 717.962 | 468.546 | 333.854 | 1537.129 | 1537.129 | 413.696 | 413.696 |
| **Overidentification Test (Outcome: Reading Scores)** | | | | | | | | |
| Hansen J Statistic | 14.301 | 16.487 | 9.341 | 15.069 | 4.622 | 3.350 | 7.652 | 5.631 |
| Hansen Test p Value | 0.112 | 0.157 | 0.674 | 0.238 | 0.706 | 0.851 | 0.239 | 0.352 |
| **Overidentification Test (Outcome: Math Scores)** | | | | | | | | |
| Hansen J Statistic | 16.921 | 10.460 | 11.177 | 9.454 | 3.666 | 3.519 | 7.865 | 6.784 |
| Hansen Test p Value | 0.257 | 0.315 | 0.514 | 0.664 | 0.817 | 0.833 | 0.312 | 0.432 |
| **N** | 10521 | 8757 | 5525 | 4625 | 7822 | 7822 | 4082 | 4082 |
| **Adj. R-Squared** | 0.852 | 0.844 | 0.703 | 0.67 | 0.846 | 0.846 | 0.787 | 0.787 |

Note: 1. Robust standard errors are in parentheses. ***, **, and * indicate the estimates are statistically significant at the 1%, 5%, and 10% levels, respectively.
2. All models include variables in control set D. For its definitions, refer to Table (2) notes for details.
3. Breusch-Pagan test is performed to examine the existence of the heteroskedasticity of the residuals in the first-stage. As shown in this table, we reject the null of no heteroscedasticity for all variables together at $p \leq 0.0001$ level.
4. Joint F test on excluded instruments is conducted and test statistic is reported. We are able to reject the null hypothesis of weak instruments when F statistic is greater than 10 (for details, see Staiger and Stock, 1997).
5. Kleibergen-Paap Wald rk F statistic for weak identification is reported (for test details, see Stock and Yogo, 2005). According to Stock-Yogo critical values, for all estimations, we have rejected the null hypothesis that instruments are weakly associated with the potential endogenous variable "speaking non-English languages at home" at $p \leq 0.001$ level.
6 Hansen overidentification test statistic and p-values are reported. For $p \geq 0.1$, we cannot reject the null hypothesis that the instruments are not correlated with test scores.
Source: Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K).

**Table 5. IV Results: the Effect of Speaking a Language Other Than English at Home on the Level of Test Scores and Growth of Test Scores**

| | Lewbel IV<br>(1) | County IV + Lewbel IV<br>(2) |
|---|---|---|
| **Panel A: Grade 3 Test Scores** | | |
| Reading Tests | -0.140*** | -0.149*** |
| | (0.030) | (0.035) |
| Math Tests | -0.078 | -0.097 |
| | (0.057) | (0.059) |
| N | 10521 | 5525 |
| **Panel B: Grade 5 Test Scores** | | |
| Reading Tests | -0.174*** | -0.159*** |
| | (0.039) | (0.048) |
| Math Tests | -0.053 | -0.054 |
| | (0.037) | (0.043) |
| N | 8757 | 4625 |
| **Panel C: Difference of Test Scores Between Grade 5 and Grade 3** | | |
| Reading | 0.004 | -0.082 |
| | (0.811) | (0.632) |
| Math | 0.812 | 0.741 |
| | (0.659) | (0.586) |
| N | 7822 | 4082 |
| **Panel D: Growth Rate of Test Scores from Grade 3 to Grade 5** | | |
| Reading | 0.007 | 0.006 |
| | (0.008) | (0.007) |
| Math | 0.012 | 0.014 |
| | (0.009) | (0.010) |
| N | 7822 | 4082 |
| Control Set | D | D |

Note: 1. Robust standard errors are in parentheses. The standard errors in Column 2 are clustered at county level. ***, **, and * indicate the estimates are statistically significant at the 1%, 5%, and 10% levels, respectively.
2. For definitions of control set D, please refer to Table (3) for details.
3. The dependent variable in Panel C is the difference of test scores between the third grade and the fifth grade. The dependent variable in Panel D is the growth rate of test scores from the third grade to the fifth grade.
Source: Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K).

**Table 6. Results for Specification Checks**

| | State IV + Lewbel IV (1) | State IV + Lewbel IV (2) | County IV + Lewbel IV (3) |
|---|---|---|---|
| **Specification** | Use State IV | Restrict to Large Counties | Exclude Non-U.S. Born |
| **Panel A: Grade 3 Test Scores** | | | |
| Reading Tests | -0.133*** | -0.153*** | -0.159*** |
| | (0.031) | (0.034) | (0.058) |
| Math Tests | -0.071 | -0.098 | -0.104 |
| | (0.047) | (0.064) | (0.059) |
| N | 10521 | 5525 | 5388 |
| **Panel B: Grade 5 Test Scores** | | | |
| Reading Tests | -0.185*** | -0.167*** | -0.186*** |
| | (0.039) | (0.046) | (0.062) |
| Math Tests | -0.047 | -0.056 | -0.056 |
| | (0.038) | (0.044) | (0.045) |
| N | 8757 | 4625 | 4489 |
| **Panel C: Difference of Test Scores Between Grade 5 and Grade 3** | | | |
| Reading Tests | 0.002 | -0.049 | -0.272 |
| | (0.644) | (0.634) | (1.028) |
| Math Tests | 0.843 | 0.656 | 0.657 |
| | (0.528) | (0.592) | (0.797) |
| N | 7822 | 4082 | 3875 |
| **Panel D: Growth Rate of Test Scores from Grade 3 to Grade 5** | | | |
| Reading Tests | 0.007 | 0.007 | 0.005 |
| | (0.006) | (0.007) | (0.011) |
| Math Tests | 0.012 | 0.013 | 0.014 |
| | (0.009) | (0.010) | (0.012) |
| N | 7822 | 4082 | 3975 |
| Control Set | D | D | D |

Note: 1. Robust standard errors are in parentheses. The standard errors in Column 2 are clustered at county level. ***, **, and * indicate the estimates are statistically significant at the 1%, 5%, and 10% levels, respectively.
2. For definitions of control set D, please refer to Table (3) for details.
3. The dependent variable in Panel C is the difference of test scores between the third grade and the fifth grade. The dependent variable in Panel D is the growth rate of test scores from the third grade to the fifth grade.
Source: Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K).

**Table 7. Effects of Speaking Non-English Languages at Home on Test Scores: Subgroup Analysis**

| | Lewbel IV<br>(1) | County IV + Lewbel IV<br>(2) | Lewbel IV<br>(3) | County IV + Lewbel IV<br>(4) | Lewbel IV<br>(5) | County IV + Lewbel IV<br>(6) | Lewbel IV<br>(7) | County IV + Lewbel IV<br>(8) |
|---|---|---|---|---|---|---|---|---|
| | **Spanish or English** | | **Non-Spanish** | | **Boys** | | **Girls** | |
| **Panel A: Grade 3 Test Scores** | | | | | | | | |
| Reading | -0.139 | -0.143*** | -0.108*** | -0.112*** | -0.150*** | -0.161*** | -0.139*** | -0.147*** |
| | (0.046) | (0.031) | (0.047) | (0.045) | (0.057) | (0.059) | (0.057) | (0.065) |
| Math | -0.047 | -0.053 | -0.033 | -0.018 | -0.048 | -0.064 | -0.113*** | -0.121*** |
| | (0.047) | (0.047) | (0.049) | (0.068) | (0.061) | (0.075) | (0.059) | (0.054) |
| N | 10152 | 5249 | 10114 | 5229 | 5335 | 2788 | 5186 | 2737 |
| **Panel B: Grade 5 Test Scores** | | | | | | | | |
| Reading | -0.187 | -0.118*** | -0.153*** | -0.110* | -0.196 | -0.164** | -0.201*** | -0.148*** |
| | (0.047) | (0.044) | (0.052) | (0.061) | (0.063) | (0.068) | (0.061) | (0.064) |
| Math | -0.039 | 0.009 | -0.013 | 0.012 | 0.023 | 0.052 | -0.153*** | -0.138*** |
| | (0.051) | (0.041) | (0.055) | (0.078) | (0.066) | (0.074) | (0.066) | (0.055) |
| N | 8411 | 4360 | 8368 | 4318 | 4393 | 2289 | 4364 | 2336 |
| **Control Sets** | D | D | D | D | D | D | D | D |

Note: 1. Robust standard errors are in parentheses. The standard errors in Column 2 are clustered at county level. ***, **, and * indicate the estimates are statistically significant at the 1%, 5%, and 10% levels, respectively.
2. For definitions of control set D, please refer to Table (3) for details.
Source: Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K).

## Appendix Table 1. Summary Statistics (Control Variables)

| Variables | Set | Language Environment at Home<br>English<br>(1) | Non-English<br>(2) | Difference<br>(3)=(1)-(2) |
|---|---|---|---|---|
| | | **Grade 5 Sample** | | |
| **Race/Ethnicity** | | | | |
| White | A | 0.71 | 0.07 | 0.64*** |
| | | (0.45) | (0.25) | |
| Black | A | 0.11 | 0.01 | 0.1*** |
| | | (0.31) | (0.11) | |
| Hispanic | A | 0.10 | 0.62 | -0.52*** |
| | | (0.30) | (0.49) | |
| Asian | A | 0.03 | 0.27 | -0.24*** |
| | | (0.16) | (0.44) | |
| Other | A | 0.06 | 0.03 | 0.02*** |
| | | (0.23) | (0.18) | |
| **Male** | A | 0.50 | 0.49 | 0.01 |
| | | (0.50) | (0.50) | |
| **Age (in months)** | | | | |
| 148-162 months | A | 0.01 | 0.02 | -0.01 |
| | | (0.10) | (0.12) | |
| 163-168 months | A | 0.29 | 0.33 | -0.04*** |
| | | (0.45) | (0.47) | |
| 169-174 months | A | 0.46 | 0.48 | -0.03 |
| | | (0.50) | (0.50) | |
| 175-180 months | A | 0.23 | 0.16 | 0.06*** |
| | | (0.42) | (0.37) | |
| 181-203 months | A | 0.02 | 0.01 | 0.01*** |
| | | (0.14) | (0.09) | |
| **Birth Weight** | A | 7.00 | 6.86 | 0.15*** |
| | | (1.33) | (1.29) | |
| **Father's Education** | | | | |
| Less than High School | B | 0.05 | 0.27 | -0.22*** |
| | | (0.22) | (0.44) | |
| High School | B | 0.21 | 0.21 | 0.00 |
| | | (0.41) | (0.41) | |
| Some College | B | 0.24 | 0.17 | 0.07*** |
| | | (0.43) | (0.37) | |
| Bachelor and Above | B | 0.31 | 0.21 | 0.10*** |
| | | (0.46) | (0.41) | |
| Education Missing | B | 0.20 | 0.15 | 0.05*** |
| | | (0.40) | (0.35) | |
| **Mother's Education** | | | | |
| Less than High School | B | 0.05 | 0.29 | -0.24*** |
| | | (0.22) | (0.45) | |
| High School | B | 0.24 | 0.25 | 0.00 |
| | | (0.43) | (0.43) | |
| Some College | B | 0.38 | 0.24 | 0.14*** |
| | | (0.49) | (0.43) | |
| Bachelor and Above | B | 0.33 | 0.22 | 0.11*** |
| | | (0.47) | (0.41) | |
| **Log Family Income** | B | 10.64 | 10.03 | 0.61*** |
| **(in 1998)** | | (1.52) | (1.73) | |

## Cont. Appendix Table 1. Summary Statistics (Control Variables)

| Variables | Set | Language Environment at Home<br>English<br>(1) | <br>Non-English<br>(2) | Difference<br>(3)=(1)-(2) |
|---|---|---|---|---|
| | | **Grade 5 Sample** | | |
| **Numbers of Books at Home** | B | 120.75 | 56.11 | 64.65*** |
| | | (183.18) | (88.82) | |
| **Own a computer at Home** | B | 0.89 | 0.77 | 0.13*** |
| | | (0.31) | (0.42) | |
| **Number of Siblings** | B | 1.51 | 1.91 | -0.4*** |
| | | (1.07) | (1.35) | |
| **Single Parents** | B | 0.20 | 0.15 | 0.05*** |
| | | (0.40) | (0.35) | |
| **Public School** | C | 0.78 | 0.91 | -0.13*** |
| | | (0.41) | (0.29) | |
| **Transfer Student** | C | 0.18 | 0.20 | -0.02*** |
| | | (0.38) | (0.40) | |
| **Region** | | | | |
| Northeast | C | 0.20 | 0.16 | 0.04*** |
| | | (0.40) | (0.37) | |
| Midwest | C | 0.29 | 0.15 | 0.15*** |
| | | (0.46) | (0.35) | |
| South | C | 0.32 | 0.24 | 0.08*** |
| | | (0.47) | (0.42) | |
| West | C | 0.18 | 0.46 | -0.27*** |
| | | (0.39) | (0.50) | |
| **Urbanicity** | | | | |
| Central City | C | 0.33 | 0.58 | -0.25*** |
| | | (0.47) | (0.49) | |
| Urban and Large Town | C | 0.40 | 0.34 | 0.06*** |
| | | (0.49) | (0.47) | |
| Small Town and Rural | C | 0.27 | 0.08 | 0.19*** |
| | | (0.44) | (0.27) | |
| **Born in the U.S.** | D | 0.99 | 0.88 | 0.11*** |
| | | (0.10) | (0.32) | |
| **Came to the U.S. After Age 3** | D | 0.00 | 0.06 | -0.05*** |
| | | (0.06) | (0.23) | |
| **N** | | 7936 | 1157 | |

Note: Standard deviations are in parentheses. ***, **, and * indicate the estimates are statistically significant at the 1%, 5%, and 10% levels, respectively. Control set A includes all variables listed in set A. Control set B includes variables in control set A and variables listed in set B. Control Set C includes variables in control set B and variables listed in set C. Control set D includes variables in control set C and variables listed in set D.
Source: Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K).

## Appendix Table 2. Full First-Stage Results of Lewbel IV Estimation

| Variables | Lewbel IV (1) | | Lewbel IV (2) |
|---|---|---|---|
| | **Grade 3** | | **Grade 5** |
| White | -0.017*** | White | -0.011* |
| | (0.005) | | (0.006) |
| Black | 0.376*** | Black | 0.396*** |
| | (0.005) | | (0.006) |
| Hispanic | 0.505*** | Hispanic | 0.514*** |
| | (0.006) | | (0.006) |
| Asian | 0.047*** | Asian | 0.047*** |
| | (0.004) | | (0.006) |
| Other | 0.001 | Other | -0.001 |
| | (0.002) | | (0.003) |
| Age | | Age | |
| 105-107 months | -0.019*** | 163-168 months | -0.014 |
| | (0.006) | | (0.017) |
| 108-110 months | -0.016** | 169-174 months | -0.013 |
| | (0.006) | | (0.017) |
| 111-113 months | -0.014 | 175-180 months | -0.015 |
| | -0.016 | | (0.017) |
| 114-116 months | -0.021 | 181-203 months | -0.021 |
| | -0.016 | | (0.020) |
| 117 or more months | -0.021 | | |
| | (0.107) | | |
| Birth Weight | 0.003*** | Birth Weight | 0.004*** |
| | (0.001) | | (0.001) |
| Dad's Education | | Dad's Education | |
| High School | -0.059*** | High School | -0.065*** |
| | (0.006) | | (0.007) |
| Some College | -0.061*** | Some College | -0.071*** |
| | (0.006) | | (0.007) |
| Bachelor and Above | -0.055*** | Bachelor and Above | -0.065*** |
| | (0.007) | | (0.008) |
| Education Missing | -0.059*** | Education Missing | -0.069*** |
| | (0.007) | | (0.007) |
| Mom's Education | | Mom's Education | |
| High School | -0.075*** | High School | -0.085*** |
| | (0.007) | | (0.008) |
| Some College | -0.071*** | Some College | -0.080*** |
| | (0.007) | | (0.008) |
| Bachelor and Above | -0.065*** | Bachelor and Above | -0.073*** |
| | (0.008) | | (0.009) |
| Ln(Family Income) in 1998 | -0.003*** | Ln(Family Income) in 1998 | -0.003** |
| | (0.001) | | (0.001) |
| Number of Books | -0.000 | Number of Books | -0.000** |
| | (0.000) | | (0.000) |
| Own a Computer | -0.011*** | Own a Computer | -0.006 |
| | (0.004) | | (0.005) |
| Number of Siblings | 0.005*** | Number of Siblings | 0.009*** |
| | (0.001) | | (0.002) |

**Cont. Appendix Table 2. Full First-Stage Results of Lewbel IV Estimation**

| Variables | Lewbel IV (1) | | Lewbel IV (2) |
|---|---|---|---|
| | **Grade 3** | | **Grade 5** |
| Public School | 0.001 | Public School | 0.004 |
| | (0.003) | | (0.004) |
| Transfer Student | 0.002 | Transfer Student | 0.006 |
| | (0.003) | | (0.004) |
| Midwest | -0.005 | Midwest | -0.005 |
| | (0.004) | | (0.005) |
| South | -0.013*** | South | -0.016*** |
| | (0.004) | | (0.004) |
| West | 0.001 | West | 0.000 |
| | (0.004) | | (0.005) |
| Urban and Large Town | -0.013*** | Urban and Large Town | -0.016*** |
| | (0.003) | | (0.004) |
| Small Town and Rural | -0.022*** | Small Town and Rural | -0.028*** |
| | (0.004) | | (0.005) |
| Born in the U.S. | -0.210*** | Born in the U.S. | -0.213*** |
| | (0.028) | | (0.027) |
| Entered in the U.S. After Age 3 | 0.144*** | Entered in the U.S. After Age 3 | 0.150*** |
| | (0.036) | | (0.037) |
| IV1(Hispanic) | 1.182*** | IV1(Hispanic) | 1.178*** |
| | (0.025) | | (0.028) |
| IV2(Asian) | 1.175*** | IV2(Asian) | 1.173*** |
| | (0.024) | | (0.028) |
| IV3(Other) | 1.117*** | IV3(Other) | 1.085*** |
| | (0.057) | | (0.070) |
| IV4(ln Income) | -0.014* | IV4(ln Income) | -0.016** |
| | (0.007) | | (0.008) |
| IV5(Midwest) | -0.139 | | |
| | (0.086) | | |
| IV6(South) | -0.162*** | IV5(South) | -0.163** |
| | (0.063) | | (0.068) |
| IV7(West) | -0.040 | | |
| | (0.052) | | |
| IV8(Small Town) | -0.300*** | IV6(Small Town) | -0.346*** |
| | (0.076) | | (0.080) |
| IV9(Entered in the U.S. After Age 3) | 0.315*** | IV7(Entered in the U.S. After Age 3) | 0.418*** |
| | (0.087) | | (0.110) |
| | | IV8(Own a Computer) | -0.041 |
| | | | (0.021) |
| Constant | 0.392*** | | 0.403*** |
| | (0.034) | | (0.038) |
| N | 8757 | | 8757 |

Note: Robust standard errors are in parentheses. The standard errors in Column 2 are clustered at county level. ***, **, and * indicate the estimates are statistically significant at the 1%, 5%, and 10% levels, respectively. Each instrument is constructed using $[Z_i - E(Z_i)] * e_i$.
Source: Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K).

## Appendix Table 3. Full First-Stage Results of County IV+ Lewbel IV Estimation

| | County IV +Lewbel IV (1) | | County IV + Lewbel IV (2) |
|---|---|---|---|
| | **Grade 3** | | **Grade 5** |
| White | -0.022* | White | -0.022 |
| | (0.012) | | (0.016) |
| Black | 0.354*** | Black | 0.374*** |
| | (0.016) | | (0.018) |
| Hispanic | 0.522*** | Hispanic | 0.517*** |
| | (0.016) | | (0.016) |
| Asian | 0.025 | Asian | 0.035* |
| | (0.017) | | (0.019) |
| Other | 0.004 | Other | 0.005 |
| | (0.004) | | (0.005) |
| Age | | Age | |
| 105-107 months | -0.009 | 105-107 months | -0.008 |
| | (0.008) | | (0.022) |
| 108-110 months | -0.008 | 108-110 months | -0.000 |
| | (0.008) | | (0.021) |
| 111-113 months | -0.005 | 111-113 months | -0.004 |
| | (0.008) | | (0.022) |
| 114-116 months | -0.012 | 114-116 months | -0.033 |
| | (0.008) | | (0.028) |
| 117 or more months | -0.014 | 117 or more months | |
| | (0.013) | | |
| Birth Weight | 0.003** | Birth Weight | 0.004** |
| | (0.001) | | (0.002) |
| Dad's Education | | Dad's Education | |
| High School | -0.110*** | High School | -0.119*** |
| | (0.013) | | (0.013) |
| Some College | -0.109*** | Some College | -0.130*** |
| | (0.012) | | (0.012) |
| Bachelor and Above | -0.097*** | Bachelor and Above | -0.117*** |
| | (0.015) | | (0.015) |
| Education Missing | -0.118*** | Education Missing | -0.138*** |
| | (0.014) | | (0.013) |
| Mom's Education | | Mom's Education | |
| High School | -0.122*** | High School | -0.146*** |
| | (0.018) | | (0.019) |
| Some College | -0.123*** | Some College | -0.140*** |
| | (0.019) | | (0.020) |
| Bachelor and Above | -0.110*** | Bachelor and Above | -0.123*** |
| | (0.017) | | (0.020) |
| Ln(Family Income) in 1998 | -0.009** | Ln(Family Income) in 1998 | -0.014*** |
| | (0.004) | | (0.005) |
| Number of Books | -0.000 | Number of Books | -0.000** |
| | (0.000) | | (0.000) |
| Own a Computer | -0.014** | Own a Computer | 0.011 |
| | (0.007) | | (0.008) |
| Number of Siblings | 0.005* | Number of Siblings | 0.008*** |
| | (0.003) | | (0.003) |

**Cont. Appendix Table 3. Full First-Stage Results of County IV+ Lewbel IV Estimation**

| | County IV +Lewbel IV (1) | | County IV + Lewbel IV (2) |
|---|---|---|---|
| | **Grade 3** | | **Grade 5** |
| Public School | 0.001 | Public School | 0.005 |
| | (0.009) | | (0.011) |
| Transfer Student | -0.004 | Transfer Student | 0.017* |
| | (0.006) | | (0.009) |
| Midwest | 0.000 | Midwest | -0.001 |
| | (0.007) | | (0.008) |
| South | -0.010 | South | -0.019** |
| | (0.007) | | (0.009) |
| West | -0.007 | West | -0.009 |
| | (0.006) | | (0.009) |
| Urban and Large Town | -0.004 | Urban and Large Town | -0.010 |
| | (0.006) | | (0.007) |
| Small Town and Rural | -0.022** | Small Town and Rural | -0.016 |
| | (0.011) | | (0.017) |
| Born in the U.S. | -0.276*** | Born in the U.S. | -0.256*** |
| | (0.030) | | (0.021) |
| Entered in the U.S. After Age 3 | 0.035 | Entered in the U.S. After Age 3 | 0.091** |
| | (0.047) | | (0.037) |
| IV1(Hispanic) | 1.260*** | IV1(Hispanic) | 1.293*** |
| | (0.076) | | (0.072) |
| IV2(Asian) | 1.214*** | IV2(Asian) | 1.251*** |
| | (0.053) | | (0.059) |
| IV3(Other) | 0.877*** | IV3(Other) | 0.879*** |
| | (0.235) | | (0.236) |
| IV4(Dad's Education High School) | -0.017 | IV4(Dad's Education High School) | 0.073* |
| | (0.048) | | (0.042) |
| IV5(Dad's Education Bachelor and Above) | 0.118 | IV5(Dad's Education Bachelor and Above) | 0.136* |
| | (0.081) | | (0.068) |
| IV6(Ln Income) | -0.027* | IV6(Ln Income) | -0.028* |
| | (0.016) | | (0.015) |
| IV7(Midwest) | -0.275 | | |
| | (0.230) | | |
| IV8(South) | -0.287* | | |
| | (0.148) | | |
| IV9(West) | -0.061 | | |
| | (0.136) | | |
| IV10(Born in the U.S. ) | -0.334*** | IV7(Born in the U.S. ) | -0.250*** |
| | (0.094) | | (0.079) |
| IV11(Entered in the U.S. After Age 3) | 0.062*** | IV8(Entered in the U.S. After Age 3) | 0.234*** |
| | (0.112) | | (0.135) |
| IV12(% Population Speaking Non-English Languages) | 0.116*** | IV9(% Population Speaking Non-English Languages) | 0.132*** |
| | (0.027) | | (0.031) |
| Constant | 0.578*** | | 0.623*** |
| | (0.068) | | (0.072) |
| N | 4625 | | 4625 |

Note: Note: Robust standard errors are in parentheses. The standard errors in Column 2 are clustered at county level. ***, **, and * indicate the estimates are statistically significant at the 1%, 5%, and 10% levels, respectively. Each instrument is constructed using $[Z_i - E(Z_i)] * e_i$

Source: Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K).

## Appendix Table 4. Diagnostic Tests of Instruments in IV Estimation: Sample of Spanish & Non-Spanish

| | **Lewbel IV** | **Lewbel IV** | **County IV +Lewbel IV** | **County IV + Lewbel IV** | **Lewbel IV** | **Lewbel IV** | **County IV +Lewbel IV** | **County IV + Lewbel IV** |
|---|---|---|---|---|---|---|---|---|
| | **(1)** | **(2)** | **(3)** | **(4)** | **(5)** | **(6)** | **(7)** | **(8)** |
| | **Speaking Spanish or English** | | | | **Speaking Non-Spanish** | | | |
| | **Grade 3** | **Grade 5** | **Grade 3** | **Grade 5** | **Grade 3** | **Grade 5** | **Grade 3** | **Grade 5** |
| **Coefficients on External Instruments** | | | | | | | | |
| % Population Speaking Non-English at Home | | | 0.037*** | 0.049*** | | | 0.101*** | 0.094*** |
| | | | (0.013) | (0.013) | | | (0.034) | (0.026) |
| **Heteroskedasticity Test** | | | | | | | | |
| Breusch-Pagan Test Stat. | 13487.531 | 10149.352 | 4720.206 | 3289.734 | 11951.162 | 8376.061 | 4213.043 | 2871.617 |
| Breusch-Pagan p Value | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **Weak Instruments Tests** | | | | | | | | |
| F Stat. on Instruments | 3878.000 | 3307.041 | 1385.283 | 1733.914 | 2076.770 | 1515.667 | 543.254 | 498.140 |
| Kleibergen-Paap F Stat. | 3578.156 | 3479.210 | 1149.052 | 1846.138 | 1931.751 | 1632.951 | 493.312 | 486.393 |
| **Overidentification Test (Outcome: Reading Scores)** | | | | | | | | |
| Hansen J Statistic | 4.643 | 10.435 | 10.031 | 12.611 | 6.186 | 12.989 | 4.972 | 13.890 |
| Hansen Test p Value | 0.703 | 0.165 | 0.187 | 0.131 | 0.518 | 0.172 | 0.663 | 0.153 |
| **Overidentification Test (Outcome: Math Scores)** | | | | | | | | |
| Hansen J Statistic | 10.816 | 9.913 | 8.975 | 13.892 | 13.663 | 10.010 | 5.585 | 11.289 |
| Hansen Test p Value | 0.147 | 0.194 | 0.254 | 0.125 | 0.158 | 0.188 | 0.589 | 0.126 |
| Control Set | D | D | D | D | D | D | D | D |
| N | 101520 | 8411 | 5249 | 4360 | 10114 | 8368 | 5229 | 4318 |
| Adj. R-Squared | 0.906 | 0.909 | 0.864 | 0.832 | 0.807 | 0.79 | 0.631 | 0.611 |

Note: 1. Robust standard errors are in parentheses. ***, **, and * indicate the estimates are statistically significant at the 1%, 5%, and 10% levels, respectively.
2. For definitions of control set D, refer to Table (3) for details.
3. Breusch-Pagan test is performed to examine the existence of the heteroskedasticity of the residuals in the first-stage. As shown in this table, we reject the null of no heteroscedasticity for all variables together at $p \leq 0.0001$ level.
4. Joint F test on excluded instruments is conducted and test statistic is reported. We are able to reject the null hypothesis of weak instruments when F statistic is greater than 10 (for details, see Staiger and Stock, 1997).
5. Kleibergen-Paap Wald rk F statistic for weak identification is reported (for test details, see Stock and Yogo, 2005). According to Stock-Yogo critical values, for all estimations, we have rejected the null hypothesis that instruments are weakly associated with the potential endogenous variable "speaking non-English languages at home" at $p \leq 0.001$ level.
6 Hansen overidentification test statistic and p-values are reported. For $p \geq 0.1$, we cannot reject the null hypothesis that the instruments are not correlated with test scores.
Source: Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K).

## Appendix Table 5. Diagnostic Tests of Instruments in IV Estimation: Sample of Boys & Girls

| | Lewbel IV | Lewbel IV | County IV +Lewbel IV | County IV +Lewbel IV | Lewbel IV | Lewbel IV | County IV +Lewbel IV | County IV +Lewbel IV |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| | Boys | | | | Girls | | | |
| | Grade 3 | Grade 5 | Grade 3 | Grade 5 | Grade 3 | Grade 5 | Grade 3 | Grade 5 |
| **Coefficients on External Instruments** | | | | | | | | |
| % Population Speaking | | | 0.116*** | 0.138*** | | | 0.124*** | 0.135*** |
| Non-English at Home | | | (0.035) | (0.039) | | | (0.034) | (0.037) |
| **Heteroskedasticity Test** | | | | | | | | |
| Breusch-Pagan Test Stat. | 4819.877 | 3486.085 | 1665.220 | 1113.904 | 4686.516 | 3369.319 | 1591.387 | 1113.175 |
| Breusch-Pagan p Value | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **Weak Instruments Tests** | | | | | | | | |
| F Stat. on Instruments | 1070.355 | 789.143 | 292.757 | 213.958 | 1358.818 | 997.901 | 748.351 | 531.293 |
| Kleibergen-Paap F Stat. | 910.129 | 964.253 | 278.093 | 210.459 | 1218.963 | 857.301 | 484.035 | 352.154 |
| **Overidentification Test (Outcome: Reading Scores)** | | | | | | | | |
| Hansen J Statistic | 7.199 | 7.742 | 8.922 | 10.849 | 5.389 | 13.035 | 12.159 | 15.614 |
| Hansen Test p Value | 0.408 | 0.356 | 0.710 | 0.542 | 0.613 | 0.071 | 0.433 | 0.210 |
| **Overidentification Test (Outcome: Math Scores)** | | | | | | | | |
| Hansen J Statistic | 12.583 | 7.097 | 7.207 | 11.998 | 12.451 | 10.362 | 13.141 | 12.585 |
| Hansen Test p Value | 0.183 | 0.419 | 0.844 | 0.446 | 0.117 | 0.169 | 0.359 | 0.400 |
| Control Set | D | D | D | D | D | D | D | D |
| N | 5335 | 4393 | 2788 | 2289 | 5186 | 4364 | 2737 | 2336 |
| Adj. R-Squared | 0.841 | 0.833 | 0.803 | 0.789 | 0.848 | 0.838 | 0.714 | 0.681 |

Note: 1. Robust standard errors are in parentheses. ***, **, and * indicate the estimates are statistically significant at the 1%, 5%, and 10% levels, respectively.
2. For definitions of control set D, refer to Table (2) notes for details.
3. Breusch-Pagan test is performed to examine the existence of the heteroskedasticity of the residuals in the first-stage. As shown in this table, we reject the null of no heteroscedasticity for all variables together at $p \leq 0.0001$ level.
4. Joint F test on excluded instruments is conducted and test statistic is reported. We are able to reject the null hypothesis of weak instruments when F statistic is greater than 10 (for details, see Staiger and Stock, 1997).
5. Kleibergen-Paap Wald rk F statistic for weak identification is reported (for test details, see Stock and Yogo, 2005). According to Stock-Yogo critical values, for all estimations, we have rejected the null hypothesis that instruments are weakly associated with the potential endogenous variable "speaking non-English languages at home" at $p \leq 0.001$ level.
6 Hansen overidentification test statistic and p-values are reported. For $p \geq 0.1$, we cannot reject the null hypothesis that the instruments are not correlated with test scores.
Source: Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K)

## APPENDIX
## IRB APPROVAL LETTER FOR CHAPTER 3

University *of* New Hampshire

Research Integrity Services, Service Building
51 College Road, Durham, NH 03824-3585
Fax: 603-862-3564

27-Sep-2013

Gao, Jia
Economics, Paul College
Durham, NH 03824

**IRB #:** 5828
**Study:** The Determinants of Students' Cognitive and Non-Cognitive Achievement During Early Childhood
**Approval Date:** 24-Sep-2013

The Institutional Review Board for the Protection of Human Subjects in Research (IRB) has reviewed and approved the protocol for your study as Exempt as described in Title 45, Code of Federal Regulations (CFR), Part 46, Subsection 101(b). Approval is granted to conduct your study as described in your protocol.

Researchers who conduct studies involving human subjects have responsibilities as outlined in the attached document, *Responsibilities of Directors of Research Studies Involving Human Subjects.* (This document is also available at http://unh.edu/research/irb-application-resources.) Please read this document carefully before commencing your work involving human subjects.

Upon completion of your study, please complete the enclosed Exempt Study Final Report form and return it to this office along with a report of your findings.

If you have questions or concerns about your study or this approval, please feel free to contact me at 603-862-2003 or Julie.simpson@unh.edu. Please refer to the IRB # above in all correspondence related to this study. The IRB wishes you success with your research.

For the IRB,

Julie F. Simpson
Director

cc: File
Houtenville, Andrew