ESSAYS IN HEALTH ECONOMICS AND LABOR ECONOMICS

by

Philip Paul DeCicca

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in The University of Michigan
2005

Doctoral Committee:

Professor Charles C. Brown, Co-Chair
Professor John E. DiNardo, Co-Chair
Professor Michael E. Chernew
Associate Professor Kerwin K. Charles

To my wife Jessica,
and my parents Richard and Vivian

ACKNOWLEDGEMENTS

First and foremost I thank my wife, Jessica Farrell, for all of her love, patience and support over the past five years. It is an understatement to say that I would never have accomplished this without her. In every way imaginable, she is my greatest gift. I hope that I am able to return such selfless love to her for the rest of my life.

I thank my family for their patience, love and support in this endeavor and throughout my life. My parents, Richard and Vivian DeCicca made many sacrifices so that I could have opportunities beyond their own. I am truly blessed to have them as parents. As a teacher, my aunt, June Brown, introduced me to learning at an early age. More recently, she was always there to listen to me and provide words of encouragement in difficult times. My brothers, Rich and Mike DeCicca, likewise always provided me with words of encouragement and support when the going was tough. Beyond raising a wonderful daughter (for which I am eternally thankful), I thank my parents-in-law, Ed and Susan Farrell for their support and interest in my work. I owe all of these people more than I can ever repay and love them with all of my heart.

I thank my committee—Charlie Brown, John DiNardo, Kerwin Charles and Mike Chernew—for the generosity of their time, the helpfulness of their thoughts and, most importantly, their friendship. I hope that it goes without saying that my work was much improved by their insights. Other individuals I encountered at Michigan to whom I am indebted include: Scott Adams, John Bound, Mary Braun, Kitt Carpenter, John Cawley, Julie Cullen, Daniel Eisenberg, Rich Hirth, Rucker Johnson, Justin McCrary, Gary Solon and Jim Sullivan. Thanks so much to each of you!

I thank previous faculty members for their friendship and continued support. In particular, I thank Ken Couch, Don Kenkel, Dean Lillard, Alan Mathios, and Bill Rosen. All of these individuals have helped me tremendously over time and I count them among my closest friends. While all deserve special mention, Dean Lillard deserves extra-special mention. I would never have even thought of getting a Ph.D. if not for Dean. To make a long story short, he believed in me and then I believed in myself. I can only repay him by helping others as he has helped me.

Finally, I thank the friends that I have made since coming to Michigan. In particular, I thank Rodney Andrews, Juan Pablo Valenzuela Barros, Kerwin Charles, Haslyn Hunte, Pat Kline, Sara LaLumia, Rafael Portillo and Pavlo Prokopovych for their friendship and kindness over the past five years. I learned a lot at Michigan, but am certain that these relationships will endure long after I have forgotten everything!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This dissertation includes three separate chapters, each of which I describe briefly below. All three chapters empirically investigate how public policies and economic conditions affect the health and human capital formation of individuals.

Chapter 2 examines an unintended consequence of higher taxation of cigarettes. Over the past three decades, the fraction of adult smokers has fallen by more than one third, while the fraction of obese adults has nearly doubled. I examine one aspect of the possible substitution implied by these trends. In particular, I investigate whether recent historically-large cigarette tax increases led to weight gains among smokers. Using repeated cross-sectional data from the National Health Interview Surveys, I find that higher cigarette taxes are associated with an increase in the body mass index of female smokers, but find no similar gain for their male counterparts. Since weight gains among smokers, who weigh less than non-smokers, may not represent poorer health, I also investigate possible distributional impacts. I find increases in clinical obesity among these women, but no effect on the fraction clinically underweight. These findings are consistent with information which suggests that women are more likely than men to use cigarette smoking as a weight control device as well as my own sample-specific evidence that female smoking participation is more tax-sensitive than that of males. Most importantly, my findings raise the possibility that the aggregate health benefits from tax-induced smoking reductions may be smaller than anticipated.

Chapter 3 investigates the impact of full-day kindergarten on academic performance. Enrollment in full-day kindergarten has grown considerably over the past forty years—from roughly one-tenth to just over half of U.S. kindergartners today. Full-day kindergarten reappeared first in the 1960s as an intervention designed to help disadvantaged children "catch up" to their peers through additional schooling. More recently, it has gained popularity among non-poor parents and schools, so that children presently enrolled in full-day programs are, on average, very similar to their half-day counterparts in baseline test scores and other child, parent and school characteristics. Using longitudinal data from the Early Childhood Longitudinal Study, I estimate the impact of full-day kindergarten attendance on standardized test scores in mathematics and reading, as children progress from kindergarten to first grade. I find that full-day kindergarten has sizeable impacts on student achievement, but these estimated gains are short-lived, particularly for minority children. Given the additional expense of full-day kindergarten, information regarding the size, distribution and duration of gains should be of great interest to policymakers.

Chapter 4 examines the impact of local labor market conditions on health. Economists have devoted much attention to the impact of macroeconomic fluctuations on a variety of outcomes, including earnings and their distribution, criminal activity and human capital investment. Collectively, they have paid less attention to a possible connection to health. Using repeated cross sectional data from the National Health Interview Surveys, I estimate the relationship between local labor market conditions and several measures of health and health behaviors for a sample of men living in the fifty-eight largest metropolitan statistical areas (MSAs) in the United States. Since the effect of labor market conditions on health may depend on the extent to which one's present or prospective employment are impacted by them, I split my sample into groups whose employment prospects are potentially more and less likely to be affected by such fluctuations. In particular, I allow the effect of local labor market conditions to vary by race and education groups since previous research suggests the labor market outcomes of non-white and less educated individuals are relatively more affected by economic fluctuations. I also allow it to vary by one's potential "exposure" to labor market fluctuations, as measured by their predicted employment status. For those men least likely to be employed, I find consistent evidence of a procyclical relationship for body weight and psychological well-being, but no systematic relationship for a variety of health behaviors including cigarette smoking, heavy alcohol consumption, and various forms of physical exercise. Consistent with these findings, I present evidence that worsening labor market conditions lead to weight gains and reduced psychological well-being among African American men and lower psychological well-being among less educated males.

CHAPTER 2

ARE HEAVIER SMOKERS AN UNINTENDED CONSEQUENCE OF HIGHER CIGARETTE TAXES?

I. Introduction

Over the past three decades, the fraction of adult smokers has fallen considerably—from roughly two-fifths to under one quarter of the adult population. Over the same period, the fraction of individuals labeled as clinically obese has nearly doubled. In this paper, I examine one aspect of the possible substitution suggested by these opposing trends. In particular, I investigate whether recent cigarette tax increases led to weight gains among likely smokers.

Using data from the National Health Interview Surveys for the years 1997 to 2001, I find that higher cigarette taxes are associated with an increase in the body mass index (BMI) of likely female smokers. Since weight gains among smokers, who are more likely than non-smokers to be clinically underweight, may not represent poorer health, I investigate possible distributional impacts and find evidence of gains in the right tail of the BMI distribution. In particular, I find increases in clinical obesity among these women, but no effect on the fraction clinically underweight. I find no evidence of any systematic relationship between BMI, or any related threshold, and cigarette taxes for men.

These findings are consistent with information which suggests that women are more likely than men to use cigarette smoking as a weight control device as well as my own sample-specific evidence that female smoking participation is more tax-sensitive than that of males. Moreover, I find that those females who gain the most weight are most responsive to taxes. Taken together, this evidence provides a possible explanation for the gender difference in my estimates. Most importantly, my findings raise the possibility that the aggregate health benefits from tax-induced reductions in smoking may be smaller than anticipated.

In the next section, I provide background on why higher cigarette taxes may lead to heavier smokers and also some reasons we might care if they do. I then describe my data in detail, focusing on the construction of key variables, how taxes are assigned to individuals, and the similarity of areas that do and do not experience a tax increase over the period in question. I also describe my analysis sample, which is used to generate most estimates. Next, I present my empirical strategy which assumes that cigarette taxes have little or no impact on the weight of non-smokers. Consistent with this notion, I allow the estimated effect of cigarette taxes on various measures of body weight to vary by smoking status. However, since tax changes may affect the composition of smokers in a given area over time, I also examine the relationship between cigarette taxes and weight for likely smokers. In the final sections, I present and discuss my estimates.

II. Background

While a possible connection between the opposing trends in smoking and obesity is of considerable interest, I focus on a narrower question.[1] Namely, I investigate

[1] Lakdawalla and Philipson (2002), Cutler, Glaeser and Shapiro (2003) and Chou, Grossman and Saffer (2004) attempt to explain the upward trend in adult obesity in the U.S. Of these papers, Chou, Grossman

whether higher cigarette taxes increase the body weight of likely smokers. The idea is not new. As noted by Philipson and Posner (1999):

> Anti-smoking measures may increase obesity and by doing so reduce the health benefits of these measures because smoking is a method of weight control and so the heavy taxes and other regulations aimed at smokers may induce people to be overweight in a Pareto sense.

In the following paragraphs, I provide some possible explanations for why higher cigarette taxes might result in heavier smokers and then discuss reasons why we might care if they do.

A. Why might higher cigarette taxes lead to heavier smokers?

Beyond anecdotal evidence, several clinical studies find that smokers gain weight when they stop smoking or reduce their consumption of cigarettes (c.f., Gritz et al., 1989; Klesges et al., 1989; Perkins, 1993; French and Jeffery, 1995). On average, smoking cessation is associated with a weight gain of five to ten pounds (USDHHS, 1990; Williamson et al., 1991), but more recent evidence suggests that the appropriate figure might be closer to ten to fifteen pounds (Klesges et al., 1998). Moreover, there is evidence that, on average, women gain more weight than men and that they are also more likely to be "supergainers", which implies a weight gain of at least thirteen kilograms (Williamson et al., 1991). Estimates from these studies, however, should be considered descriptive since changes in smoking behavior are treated as exogenous events.

Nevertheless, several physiological reasons support the likelihood of an inverse relationship between smoking and body weight. For example, there is evidence that smoking increases the body's metabolic rate, the rate at which calories are burned while an individual is at rest (Kershbaum et al., 1966; Glauser et al., 1970; Wack and Rodin,

---

and Saffer (2004) is most similar to my work since it is the only one of the three that controls for cigarette prices. These authors find that cigarette prices are directly related to BMI and clinical obesity and that they are the second most important factor, generally speaking, in explaining the growth of obesity in the United States from 1984 to 1999.

1982; Hofstetter et al., 1986). While the exact mechanism is not completely understood, researchers believe that smoking raises metabolism by stimulating the central nervous system to produce catecholamines. These hormones cause the heart to beat faster and hence lead to greater resting caloric expenditure. In addition, nicotine induces thermogenesis, the process by which the body generates heat, which also causes additional calories to be expended. As a result, former smokers tend to burn 100-200 fewer calories each day once they quit. Absent any offsetting activity, this implies a weight gain of one pound within two to four weeks, if the relationship between time and reduction in caloric expenditure is linear. In addition to its role as a metabolic stimulant, nicotine is known to be a moderately effective appetite suppressant (Grunberg, 1982). Indeed, several studies find that reductions in smoking lead to additional caloric intake, though it is difficult to place the blame squarely on nicotine, or other potential appetite suppressants in cigarette smoke, since former smokers often report an improved sense of smell and taste, which may independently increase food intake (Stamford et al., 1986; Rodin, 1987; Perkins et al, 1990; Pomerleau et al., 1991; Clearman and Jones, 1991).

Beyond physiology, available evidence suggests that many smokers use cigarettes as a weight control device. In other words, individuals who know that they are prone to gain weight and/or are more concerned with body weight in general may be more likely to smoke to control their weight. This phenomenon may be especially relevant for women. For example, Klesges and Klesges (1988) find that among a group of college students, women smokers were about sixty percent more likely to report weight control as a reason for smoking, relative to their male counterparts. Pirie et al. (1991), in a broader sample, document that female smokers were more than twice as likely to cite weight concerns as a reason for continued smoking. So, if the set of individuals who use smoking as a weight control device is also price-responsive in their smoking behavior, higher taxes might plausibly lead to weight gains among current and former smokers.

Finally, higher cigarette taxes may lead to heavier smokers even if smoking behavior is orthogonal to taxes. In particular, given their magnitude, these higher taxes may have had non-trivial reductions in smokers' disposable income.[2] Such income reductions may lead smokers to substitute into cheaper, less healthy foods or more hours worked which may lead to a wide range of other substitutions that may make them more prone to weight gain.[3] Since tobacco is an addictive good and since smokers tend to have low incomes, such effects seem plausible, especially since the vast majority of smokers do not quit altogether when faced with higher taxes.

B. Why might we care?

Since first warning Americans about the dangers of cigarette smoking in 1964, nearly all U.S. surgeons general have advocated for higher taxes on cigarettes.[4] More recently, the Department of Health and Human Services, via its *Healthy People 2010* program, has recommended a combined federal and state cigarette tax of $2.00 per pack (USDHHS, 2004).[5] This recommendation is based at least partially on evidence which suggests that higher cigarette taxes are associated with reductions in premature mortality and morbidity among smokers (GAO, 1986; Warner, 1986; Harris, 1987; Chaloupka,

---

[2] For example, assuming no change in smoking behavior, a smoker who consumes one pack of cigarettes per day will end up paying $30 extra per month following a per-pack tax increase of $1.00. For a full-time worker earning minimum wage, this represents nearly four percent of pre-tax monthly earnings.
[3] See Appendix A for a particular example.
[4] Many continue to advocate for higher taxes. In February 2004, four former surgeons general proposed a comprehensive national strategy to reduce smoking in the United States, with a two-dollar increase in the federal cigarette tax as the plan's centerpiece.
[5] Presently, only Michigan, New Jersey and Rhode Island comply with this recommendation.

1989; Moore, 1996).[6] None of these studies, however, allow for the possibility that higher taxes induce weight gains among current and former smokers. These omissions are especially relevant since, according to recent government figures, obesity-related disorders likely will soon overtake smoking as the leading cause of premature death in the United States (Mokdad et al., 2004). So, while reducing smoking-related premature mortality and avoidable morbidity is a laudable goal, it may not be fully achieved if smokers substitute weight gain for cigarette consumption or otherwise gain weight when cigarette taxes increase.

Weight gain alone, however, is not cause for concern. For example, if most of any prospective gains occur in the left tail of the weight distribution, higher taxes on cigarettes may not represent poorer health and may even result in improved health. In other words, while there is much focus on clinical obesity, being underweight also has potential health consequences. So, if higher taxes reduce the fraction of smokers who are clinically underweight, this may represent an additional health benefit of higher cigarette taxes. Alternatively, if gains are concentrated in the right tail of the distribution, higher cigarette taxes may be more likely to offset some of the health gains associated with tax-induced smoking cessation or reduction since excess body weight, like smoking, is linked to premature death and avoidable morbidity. Hence, it is important to understand the possible distributional impacts of cigarette taxes.

III. Data

I use annual cross-sectional data from the National Health Interview Surveys (NHIS) for the years 1997 to 2001, inclusive. While the NHIS dates back to 1972, it was

[6] See Chapter 6 of *Reducing Tobacco Use: A Report of the Surgeon General* (USDHHS, 2000), especially pp. 355-357, for a description of these and other related studies.

redesigned in the middle 1990s, with 1997 the first wave following this revision. I use the adult sample which consists of annual surveys of thirty to thirty-five thousand individuals. However, since the data do not include state identifiers, I restrict my analysis to individuals living in Level A or "large" metropolitan statistical areas (MSAs), for whom MSA of residence is publicly available.[7] While this creates some difficulty in assigning state-level cigarette taxes, this strategy yields between fifty and fifty-five percent of the overall NHIS sample, depending on the year in question.[8] In numbers, this amounts to nearly seventeen thousand individuals per year, for a total of about eighty-four thousand individuals, before accounting for missing data on key variables. Below, I describe these variables, focusing on body weight-related measures and cigarette taxes. I then demonstrate the initial similarity of MSAs that did and did not experience tax increases from 1997 to 2001. Finally, I provide detailed information on my analysis sample, which I use to generate most estimates.

A. Body weight-related measures

NHIS data contain information on body mass index (BMI), which is, conceptually, body weight normalized by height. More precisely, it is defined as the ratio of one's weight in kilograms divided by height in meters squared. While BMI is preferred to body weight, and is a generally-accepted metric to assess weight-related health, it has certain shortcomings. First, BMI might not be a valid measure for some individuals, perhaps due to differences in body type or composition. If not, widely-used cutoffs at the upper and lower tails of the distribution, for example, may not represent similar weight-related health for such individuals. Second, BMI information in the NHIS

[7] Level A MSAs have at least one million residents. In 1997, they contained roughly 52 percent of the U.S. population. In what follows, I use the words "MSAs" and "areas" interchangeably.
[8] Below, I explain in detail how I address this difficulty.

is constructed from self-reports of height and weight, so it is subject to measurement error (Cawley, 1999). In particular, it is likely that heavier individuals tend to under-report weight while lighter individuals over-report it. As noted by Lakdawalla and Philipson (2002), such systematic misreporting may attenuate estimated coefficients rather than merely reduce their precision, as with classical measurement error in the dependent variable. Comparing average BMI in my sample to corresponding information from the fourth National Health and Nutrition Examination Survey (NHANES IV), which took physical measurements of respondents' weight and height between 1999 and 2000, suggests that under-reporting dominates. In my sample, mean BMI is 26.23, while in NHANES IV the corresponding average for individuals eighteen years old and older is 27.85, which translates into about ten pounds for an individual of average height.[9]

In addition to BMI, itself, I focus on two widely used thresholds of weight-related health—clinical measures of underweight and obesity. The former is defined as BMI of no greater than 18.5, while the latter is defined as BMI of 30 or higher. While the primary concern is that affected smokers become clinically obese, I also consider clinical underweight since, in principle, weight gains among smokers may not harm health if they occur in the left tail of the BMI distribution. In addition, I estimate models for thresholds surrounding these clinical definitions to address systematic misreporting of BMI information. In particular, I estimate models +/- 2 units of body mass index for each threshold. These additional models also provide a more complete view of what is occurring in two very different parts of the distribution.

---

[9] The latter figure is taken from Table 1 in Chou, Grossman and Saffer (2004).

B. Cigarette taxes

In the U.S., both the federal government and state governments tax cigarettes.[10] With few exceptions, these taxes are denominated in cents per pack of twenty and are built into the purchase price of cigarettes. By their nature, federal taxes are applied to all cigarette purchases regardless of location. State taxes, by contrast, apply only to cigarette purchases made within relevant state borders. As a result, the vast majority of relevant studies, including this one, use variation in state taxes to identify changes in the particular outcome of interest.[11]

While the identifying variation in cigarette taxes occurs at the state-level, I am unable to view sampled individuals' states of residence directly. NHIS data, however, include MSA identifiers for those individuals residing in large MSAs. For individuals whose MSA lies entirely within a given state, the assignment of state level taxes is straightforward. Fortunately, this is the case for most of these large MSAs—forty-four of fifty-eight lie within the borders of a single state. Of the remaining fourteen MSAs, eleven have three-quarters or more of their population in a single state, while the remaining three have a more uniform distribution of their population in two or more states.[12] To account for this overlap, I assign population-weighted averages of state-level cigarette taxes to the fourteen MSAs in question. To check its robustness, I make two

---

[10] Local governments in a handful of states also tax cigarettes. In my sample, three MSAs (Chicago, New York and Cleveland) impose such taxes at either the city or county level. However, none changed their tax from 1997 to 2001.

[11] Some studies use a measure of average cigarette price, rather than tax information. I use tax information for two reasons. First, as noted by Evans and Ringel (1999), there is less measurement error in tax, relative to price, data. Since most variation in cigarette price is driven by tax increases, there is little lost in taking this approach. Second, I use quarterly variation in taxes, which is not available for cigarette prices. Quarterly variation should result in better timing in the assignment of taxes to individuals in areas that experienced an increase in a given calendar year.

[12] The three MSAs with substantial multiple state overlap are Kansas City (KS and MO), Providence-Fall River-Warwick (RI and MA), and Washington, DC (MD, VA, DC).

modifications to my tax variable. First, since smokers residing near state borders may "shop around" on the basis of price, I assign the minimum tax of the states that comprise each of the fourteen areas in question. Second, I eliminate residents from the three most troublesome areas from my sample, estimating models with individuals from the remaining fifty-five MSAs. I also impose these two modifications simultaneously. As I show later, my main findings are not sensitive to these modifications.

Finally, my data capture a great deal of tax variation, both cross-sectionally and over time within areas. In my 1997 sample, the lowest tax was 2.5 cents per pack in Virginia and the highest was 82.5 cents in Washington state, which corresponds to the range of all fifty states in this year. Longitudinally, twenty-nine of fifty-eight MSAs experienced an increase in tax from 1997 to 2001 and the average tax increase was roughly thirty cents, which corresponds to an increase of nearly seventy-five percent relative to 1997 taxes. This represents much more within-area variation relative to the previous five years.

C. How similar are areas that did and did not experience a tax increase?

With respect to key variables, individuals in areas that did and did not experience a tax increase between 1997 and 2001 are very similar in the base year. As seen in Table 1, mean BMI in 1997 is virtually identical across these two types of MSAs—26.231 in increasing areas and 26.232 in areas that never experienced an increase.[13] Examining various portions of these BMI distributions further reveals their similarity. For example, the fraction clinically obese in increasing areas is 18.6 percent, while it is 19.2 percent in non-increasing ones. I also compare these areas on a dimension which I call "near"

[13] By gender, the corresponding figures are, respectively, 25.913 vs. 25.902 for women and 26.641 vs. 26.653 for men.

obesity. I define this as the fraction of individuals whose BMI is greater than or equal to twenty-eight, but strictly less than thirty, the cutoff for being labeled clinically obese. Table 1 shows that the fraction near obese is virtually identical (10.7 vs. 10.8 percent) across the two types of MSAs. The differences are slightly larger in the left tail of the BMI distribution. Here, while only 1.8 percent of individuals in increasing MSAs report being clinically underweight, 2.5 percent of their counterparts in non-increasing areas report likewise. With regards to those "on the cusp" of being underweight, the figures are much more similar. In particular, while 7.4 percent of individuals in increasing MSAs report having BMI strictly greater than 18.5 and less than or equal to 20.5, a similar 7.5 percent of individuals in non-increasing MSAs do likewise. More concisely, Figures 1A and 1B present kernel density estimates for log BMI in 1997 for men and women, respectively, by MSA type. In each figure, the two densities appear to correspond closely, suggesting that the baseline BMI distributions in these two types of MSAs are indeed very similar. Moreover, separate two-sample Kolmogrov-Smirnov tests fail to reject the equality of the empirical distribution functions corresponding to the kernel density estimates in Figures 1A and 1B. Exact p-values from these tests are p=0.170 for women and p=0.353 for men.

Perhaps most surprising, however, is the relatively minor difference in base year cigarette taxes. While areas that eventually experienced an increase have, as one might expect, a higher average tax, the margin is perhaps less than anticipated. In particular, increasing areas have an average tax in 1997 of just over a nickel higher than those that do not (42.5 vs. 37.0 cents). By the end of 2001, the average nominal tax for areas that experienced an increase was 73.7 cents, a nominal gain of 31.2 cents or seventy-three percent. In real terms, this represents an increase of 28.2 cents or sixty-six percent. As seen in Figure 2, there was substantial divergence in tax regimes over this period.

D. Analysis sample

Restricting my sample to those who live in large MSAs, as described above, yields 83,936 individuals from five years of data.[14] Further limiting my sample to those with valid BMI information reduces this figure to 35,634 males and 44,878 females for a total of 80,512 individuals.[15] While this represents a loss of only four percent of the original cases, it is not random with respect to gender as nearly three percent of men, but a full five percent of women failed to provide this information. A more important question, however, is whether such missingness is related to cigarette taxes. To assess this, I regress a variable that equals one if BMI information is missing, and zero otherwise, on cigarette tax and a set of MSA and year-specific quarter fixed effects for men and women separately. Estimates from these models imply no systematic relationship between missing BMI and cigarette taxes for either group. In particular, the estimated tax coefficient for men is -0.000039 (t=-0.33) and for women it is 0.000045 (t=0.38).

Finally, with respect to key variables, my analysis sample is nearly identical to similar data from the Behavioral Risk Factor Surveillance Survey (BRFSS) in 1999, the midpoint of the years included in my data. For example, twenty-three percent of members of both samples report being smokers in 1999, while the fraction clinically obese is also very similar as twenty-one percent of my sample versus twenty-percent of respondent report weight and height such that their body mass index is at least thirty. By

[14] I also restrict the sample to adults twenty-one to eighty-five years old and exclude pregnant women.
[15] Accounting for missing smoking information reduces these figures to 35,478 men and 44,708 women.

gender, twenty-five percent of my male sample members versus twenty-four percent of BRFSS males report being smokers, while the figures for women are equal at twenty-one percent. With respect to the fraction obese, twenty-one percent of females in my sample versus twenty percent of BRFSS females are classified as obese, while the corresponding figures for men are equal at twenty percent. Information on weight and height in BRFSS is also self-reported and therefore subject to the same systematic misreporting as my data.

IV. Empirical Strategy

As noted by others, unobserved heterogeneity is perhaps the most important concern in relating cigarette taxes to any outcome of interest. In the present context, the concern is that unobserved area characteristics that are correlated with cigarette tax levels, and exert an independent influence on weight, will result in biased estimates of the relationship between these taxes and body weight. For example, it is commonly believed that states with higher cigarette taxes are somehow more "progressive" or otherwise express a stronger collective preference for health. As a result, regressing weight on cigarette taxes in a single cross-section of data may bias down the true effect of taxes if residents of high tax areas are, as implied, relatively more health conscious and this healthy disposition extends to weight-related health. The repeated cross-sectional nature of NHIS data allow for inclusion of area fixed effects, which will eliminate the troublesome heterogeneity if it is time invariant.

With this in mind, a model that bases statistical identification on within-area variation in cigarette taxes is given by:

$$\mathbf{BMI}_{ijqt} = \tau \mathbf{T}_{jqt} + \rho \mathbf{S}_{ijqt} + \beta \mathbf{X}_{ijqt} + \mu_j + \alpha_{qt} + \varepsilon_{ijqt} \qquad \mathbf{(1)}$$

Here, i indexes the individual, j MSA of residence, q quarter surveyed, and t year surveyed. BMI represents log of body mass index, T cigarette tax, S smoking status, X a set of individual and MSA-specific covariates, μ is a vector of MSA fixed effects, α is a vector of year-specific quarter fixed effects and ε is an error term with mean zero.

This specification, however, has one prominent drawback. It imposes the same relationship between taxes and weight on smokers and non-smokers alike. If cigarette taxes affect the weight of any group, they should predominantly impact the weight of smokers, for whom they are most binding. This logic implies that the relationship between cigarette taxes and weight is more reasonably given by:

$$\mathbf{BMI}_{ijqt} = \tau \mathbf{T}_{jqt} + \rho \mathbf{S}_{ijqt} + \gamma(\mathbf{S}^*\mathbf{T})_{ijqt} + \beta \mathbf{X}_{ijqt} + \mu_j + \alpha_{qt} + \varepsilon_{ijqt} \qquad \mathbf{(2)}$$

Here, the effect of taxes on weight is allowed to vary by smoking status. In effect, this strategy posits non-smokers as a control group. While likely an improvement on equation (1), this specification does not account for the possibility that taxes may also, and more directly, influence smoking initiation and cessation decisions. That is, higher taxes may change who is and who is not a smoker over time.

Such compositional change is relevant since it may impact the relationship of interest. For example, if taxes induce cessation behavior among current smokers, and if such individuals do indeed gain weight, these gains will be incorrectly assigned to non-smokers as a group. As specified in equation (2), the estimated relationship between higher taxes and weight would be attenuated. On the other margin, higher taxes may deter individuals from smoking initiation. In this case, the impact on the estimated relationship between taxes and weight is ambiguous. For example, if new smokers tend to weigh less than those who do not start smoking, reduced initiation would lead the relationship to be overstated. Conversely, if those who initiate smoking tend to be heavier, this would lead the relationship to be attenuated.[16] As a practical matter, I exclude individuals less than twenty-one years old. Since most adult smokers initiate smoking prior to this age, any change in the composition of smokers is likely driven by quitting behavior. Figures 3A and 3B present the fraction of smokers over time by whether or not their MSA of residence experienced a tax increase for men and women, respectively. While the fraction of male smokers in the two MSA-types roughly parallel each other, the greater decline in smoking among women in MSAs that experienced a tax increase demonstrates the necessity of dealing with this potential compositional problem.

I address this issue by allowing the relationship between cigarette taxes and weight to differ not only by actual smoking status, but also by *predicted* smoking status. Here, the expectation is that relative to their actual smoking status continuing smokers and those induced to quit by higher taxes are more alike in their predicted smoking behavior. Moreover, current smokers and tax-induced quitters should be more similar in their predicted smoking behavior than, for example, current smokers and individuals who have never smoked. To generate predicted smoking probabilities, I first estimate a cross-sectional model of smoking participation using my initial year of data. This model is intended to capture the data generating process for smoking prior to the divergence in tax regimes over the next five years. Using estimated coefficients from this model, I compute predicted smoking probabilities for all individuals with useable smoking and MSA of residence information.[17] Next, I split this distribution into quartiles and estimate

[16] Cawley, Markowitz and Tauras (2004) find that young females who are overweight are more likely to start smoking than their non-overweight counterparts, but find no similar evidence for young males.
[17] More precisely, models that generate the predicted probabilities are linear probability models and the predicted probability is given generally by $X_{ijqt}'\beta_{97}$, where β_{97} is the vector of coefficient estimates from the

the relationship between taxes and weight separately for each quartile, as specified in equation (1). This strategy effectively divides my sample into groups whose weight is more and less likely to be impacted by cigarette tax changes. Following standard practice in the smoking and obesity literatures, I estimate models for men and women separately.

A. Selected characteristics by quartile

As seen in Table 2, the fraction of actual smokers rises monotonically with the quartiles of predicted smoking behavior and the gradient appears to be roughly similar for men and women. The quartiles also capture differences in smoking intensity as the fraction of all individuals in a given quartile who smoke at least one pack of cigarettes per day rises similarly. With respect to weight-related outcomes, body mass index and the fraction clinically obese rise sharply with quartiles for women, but not for men. The female pattern runs counter to the notion that smokers weigh less than their non-smoking counterparts, though figures in Table 2 are based on all years of data so, to some extent, this pattern could be due to tax changes. Finally, average cigarette tax falls consistently, though not dramatically, with increasing quartiles for both men and women. Coupled with the quartile-specific information on body mass index, this implies an inverse relationship between taxes and weight for women. As I discuss next, I find a direct relationship when I relate cigarette taxes to BMI using within-area variation in taxes.

V. Results

Before discussing my main results, I briefly present estimates from models that directly compare the response of smokers vs. non-smokers, without addressing the compositional issue discussed in the previous section. I then present estimates from my

---

cross-sectional model and X_{ijqt} represents the characteristics of individual i residing in MSA j in quarter q of year t. Note also that probabilities generated by logit or probit methods yield roughly the same eventual results.

primary models, which estimate the relationship between cigarette taxes and weight-related outcomes by predicted smoking quartiles. I find evidence of a positive association between cigarette taxes and body mass index for the most likely female smokers, but not for similar men. Since this finding represents an average gain, and since weight gains among some smokers may be health-improving, or at least health-neutral, I estimate models designed to provide insight on the portion of the BMI distribution that is responsible for this average effect. I find evidence of gains for these women in the right tail of the BMI distribution.

A. Estimates by actual smoking status

Since cigarette taxes should predominantly impact the weight of smokers, rather than non-smokers, a straightforward empirical strategy involves comparing the response of smokers to non-smokers within areas, as in equation (2). However, as discussed, higher cigarette taxes may influence the composition of smokers in areas that experience an increase.[18] Recall, if those who quit smoking altogether due to higher taxes do indeed gain weight, such individuals will be incorrectly treated as non-smokers and their weight gain will, in effect, be recorded on the "wrong" side of the ledger. Hence, models that directly compare smokers and non-smokers may underestimate the true impact of taxes on body mass index.

Despite this possibility, I find discernible effects of tax on log BMI for smokers relative to non-smokers, as seen in the first column of Table 3.[19] Consistent with my main estimates, which are presented below, the body mass index of both male and female smokers is related directly to cigarette taxes, though the estimated effect is considerably

[18] This logic extends to the relative magnitude of tax increases, among areas that experience one.
[19] Corresponding estimates are similar in magnitude and precision when the dependent variable is BMI rather than log BMI.

larger and more precisely estimated for females. Finally, note that the implied effect for non-smokers is, as expected, essentially zero for both males and females.

B. Estimates by predicted smoking quartile

As described earlier, I stratify individuals into smoking quartiles, based on their predicted smoking probability, in order to avoid the potential compositional problem described above. Table 4 presents estimates from regressions of log BMI on cigarette tax and a set of covariates which include education, household income relative to poverty line, age, marital status, employment status and MSA unemployment rate, as well as year-specific quarter and MSA of residence fixed effects.

Table 4 shows that the only group for whom cigarette taxes and log BMI systematically vary is females in the highest predicted smoking quartile.[20] For this group of most likely female smokers, the coefficient estimate of 0.000317 implies that a twenty-five cent tax increase leads to an average increase in BMI of about 0.21, which, for a female of average height, represents a weight gain of roughly 0.57 kilograms or 1.25 pounds.[21] The corresponding coefficient estimate for similar males is positive, but relatively small and imprecisely estimated. For these males, the implied weight gain is about one-third of one pound.[22] However, since these are average gains, they provide little information regarding where in the distribution they originate.

---

[20] In what follows, I refer frequently to individuals in the highest predicted smoking quartile as the "most likely" smokers.
[21] The average height of females in their highest predicted smoking quartile is 1.63 meters; for similar males it is 1.75 meters.
[22] Again, models with BMI, rather than log BMI, as the dependent variable yield estimates similar in magnitude and precision.

To investigate the source of these average gains, I estimate a series of linear probability models around the clinical definitions of underweight and obesity.[23] Tables 5A and 5B present estimates from the corresponding linear probability models, respectively, for men and women.[24] Each table contains six columns of estimates and is organized as follows: the first three columns correspond to thresholds centered about the clinical definition of underweight and the second three columns correspond to thresholds surrounding the clinical definition of obesity.[25] Of the first three columns, the first column represents estimates from models where the dependent variable equals one if BMI is greater than or equal to 16.5, the second column threshold is 18.5, and the third column threshold is 20.5.[26] Of the second three columns, the first column, which is fourth overall, represents models where the dependent variable equals one if BMI is greater than twenty-eight, the second column threshold is thirty and the third column threshold is thirty-two. For the sake of completeness, I include estimates from all four quartiles in each table.

Table 5A shows no systematic relationship between cigarette taxes and any of the six BMI thresholds for men, even those in the highest predicted smoking quartile. Note, however, that for those outcomes centered about clinical obesity, all coefficients for males are positive. While these estimates are consistent with higher cigarette taxes leading to right-tail increases in weight among this group of males, they are not estimated precisely enough to reject the null hypothesis of no relationship in any of these models.

---

[23] Note also that this is a strategy to deal with self-reported measures of body mass index since reported thresholds may not correspond to actual ones.

[24] Note that coefficients from linear probability models are nearly identical to probit and logit marginal effects in all cases.

[25] Recall, clinical underweight is defined as BMI of 18.5 or less, while clinical obesity is BMI of at least 30.

[26] Very few males are clinically underweight so estimates from models at or below this threshold should be viewed with caution. The BMI $\leq$ 20.5 threshold models, however, may yield more meaningful estimates.

Table 5B, which presents estimates for females, tells a different story. While there is no systematic relationship between cigarette taxes and any of the six BMI thresholds for those in the first three predicted smoking quartiles, the estimates suggest that, for the most likely female smokers, higher cigarette taxes lead to increases in the proportion obese and the proportions corresponding to the surrounding thresholds. In all three models, the relationship between taxes and the fraction in the stated category (BMI greater than or equal to twenty-eight, thirty and thirty-two) is positive and statistically different from zero at conventional levels of significance. In particular, a twenty-five cent increase in the cigarette tax is associated with a nearly two percentage point increase in the fraction of obese individuals in this upper most quartile, which represents slightly less than a seven percent gain in the overall fraction obese in the quartile.[27] The estimated relationships are slightly larger in magnitude at the two thresholds surrounding clinical obesity, as corresponding estimates imply just over an eight percent increase in the fraction of individuals with BMI greater than or equal to twenty-eight and just under an eight percent gain for the BMI threshold of thirty-two. No evidence of a systematic relationship is found for clinical underweightness and its surrounding thresholds. Hence, for those females most likely to be smokers, it appears that the average effect of taxes on log BMI is generated by weight gains in the right tail of the distribution.

C. Why women and not men?

In other words, is my finding that higher cigarette taxes induce weight gains among the most likely female smokers, but not similar males, plausible? Taken together, two additional findings may help explain the gender difference. First, males and females

---

[27] This figure is calculated off a base of twenty-seven percent, which is the fraction of those females in the highest predicted smoking quartile who are clinically obese.

may have different motivations for smoking. As noted earlier, it is well-established that women report using smoking as a weight control device more frequently than men. In my sample, quartile-specific information on subjective overweightness lends support to this notion. As seen in Table 6, while forty-nine percent of women in the lowest predicted smoking quartile view themselves as at least five percent above their *desired* body weight, nearly sixty-five percent in the highest quartile report likewise, a difference of about thirty-three percent. For men, the gradient is much flatter, with sixty-nine percent in the lowest quartile and just over seventy-one percent in the highest, a difference of only four percent. [28] Similar patterns obtain for the fraction of women and men who view themselves as at least ten and at least twenty percent above their desired weight.

Second, my estimates suggest that female smoking is more tax-responsive than that of males.[29] As seen in Table 7, the implied price elasticity of smoking participation for all women is -0.34, while the corresponding elasticity for all men is slightly positive, but roughly zero.[30] More importantly, Table 7 also contains estimates of price sensitivity by quartile. As seen in its second column, females in the highest predicted smoking quartile are much more price sensitive than their counterparts in the lower three quartiles. This is consistent with my finding of a direct relationship between cigarette taxes and weight for these women. Interestingly, females in the second quartile are the next most price sensitive group among females. Looking back at Table 4, the relationship between

[28] To avoid possible influences of cigarette taxes, these figures are based on 1997 data.

[29] In principle, it is not clear why either gender would be more price-sensitive. However, since models tend to measure full price effects (i.e., substitution *and* income effects), rather than a pure price effect, lower income may be responsible for females' greater measured price sensitivity. Consistent with this possibility, recent evidence suggests that lower income individuals tend to be relatively more price sensitive (Gruber and Koszegi, 2004).

[30] Since the consensus estimate of the adult participation elasticity seems to lie between zero and -0.25, this figure likely represents a reasonably high degree of price sensitivity (Chaloupka and Warner, 2000).

taxes and BMI is next largest in magnitude for females in this group, following those females in the highest predicted smoking quartile. In other words, the two quartiles with the greatest estimated tax-sensitivity (i.e., fourth and second) are those that exhibit the strongest relationships between taxes and weight, suggesting that taxes are indeed responsible for the estimated gains.

The picture, however, is not as clear as it might seem. While females, especially those most likely to be smokers, are more tax-sensitive than males on the extensive smoking margin, the pattern is reversed when I examine smoking on the intensive margin. Following a long literature in health economics (c.f., Duan et al., 1983), I model the typical number of cigarettes smoked per day by smokers in the same manner I modeled smoking participation. Estimates are presented in Table 8 and show that male smoking intensity is much more tax-sensitive than that of women. In particular, males in the two highest quartiles of predicted smoking are quite sensitive on the intensive margin with price elasticities of -1.129 and -0.614, respectively.[31] Nevertheless, my main analyses show no systematic effect of taxes on the weight of these, or any other, males. While this may seem at odds with the gender differences in my main estimates, research on the compensating behavior of smokers may provide some insight. Evans and Farrelly (1998), using supplemental data from two earlier National Health Interview Surveys, examine the compensating behavior of smokers in response to changes in cigarette taxes and prices. The authors find that although smokers reduce their daily consumption of cigarettes when faced with higher taxes, they also compensated by smoking longer cigarettes and those that are higher in tar and nicotine. To the extent that nicotine plays a

[31] In more practical terms, a twenty-five cent tax increase is associated with an average reduction of roughly 1.2 and 0.75 cigarettes per day, respectively, for male smokers in these two highest quartiles.

central role in the weight changes of smokers, their findings provide a possible explanation for why likely male smokers seem to experience no weight gain, despite their estimated tax-sensitivity on the intensive smoking margin.

Finally, gender differences in income may also help explain the stronger finding for women. As seen in Table 2, while twenty percent of most likely male smokers have incomes below 125 percent of the federal poverty line, just over thirty-five percent of similar females have similar incomes.[32] That is, the most likely female smokers in my sample are about seventy-five percent more likely to have very low incomes as their male counterparts. This income disparity, and the generally low incomes of the most likely female smokers, are consistent with the idea that higher cigarette taxes have income effects that may lead to weight gain (e.g., changes in the composition of healthy vs. unhealthy food), though I emphasize that I can not directly test such possibilities with these data. Lastly, while the idea of taxes having negative income shocks among the most likely female smokers may seem inconsistent with my finding that this group's smoking participation is most sensitive to taxes, it is likely that most smokers in this category experience such shocks since only a small fraction are induced to quit and this group also appears to be insensitive to taxes on the intensive margin. More generally, whether or not cigarette taxes are, on average, regressive or not, such negative income shocks almost certainly exist for some non-trivial subset of smokers.[33]

[32] In 1999, this represents an income of $10,300 for an individual and $17,350 for a family of three.

[33] Most evidence on the vertical equity of cigarette taxes is indirect since it is inferred from findings that price responsiveness falls across rising income groups. Such findings are not inconsistent with smokers experiencing negative income shocks. This is especially true since the majority of smokers do not quit altogether in response to higher taxes.

D. Sensitivity of estimates

As discussed in section III, there are fourteen MSAs which presented difficulty in assigning cigarette taxes to their residents since they encompassed two or more states. Recall that of these fourteen MSAs, eleven had at least three-quarters of their population residing in one state, while residents of the remaining three were much more uniformly distributed over the states involved. I make two modifications to my tax variable to check that assignment of taxes is not driving the estimates presented. First, I estimate models without residents of the three most troublesome MSAs since population-weighted averages are unlikely to represent the true tax regimes these individuals face. The top panel of Table 9 reports estimates for the most likely male and female smokers only. These estimates are very similar to those in the fourth rows of Tables 5A and 5B. Second, I assign residents of the fourteen MSAs with multiple state overlap the *minimum* of the state taxes involved on the assumption that smokers within such an area may "shop around" on the basis of price. Again, as seen in the middle panel of Table 9, I find only small differences relative to my main estimates for the most likely male and female smokers. For these females, tax coefficients in the log of body mass index and obesity models are somewhat larger and more precisely estimated. The similarity of these two sets of estimates is perhaps not too surprising given that neighboring states tend to have similar cigarette tax rates. Finally, in the bottom panel of Table 9, I combine both strategies and again find very similar, if somewhat stronger, results. It is possible that the somewhat larger and more precisely estimated gains in these alternative models are due to reduction in measurement error induced by my original tax-assignment strategy.

E. Are these magnitudes plausible?

In particular, is my finding that a twenty-five cent tax increase leads to an average gain of 1.25 pounds among most likely female smokers plausible? An additional piece of information may be useful in this regard. Among most likely female smokers, just over five percent of this group reported quitting smoking within one year of being surveyed. In numbers, this translates into 575 "recent" quitters. The estimated average gain of 1.25 pounds, in response to a twenty-five cent tax increase, corresponds to an aggregate weight gain of roughly 14,100 pounds for this group. So, if only the weight of recent quitters was impacted by higher taxes, each would have to gain nearly twenty-five pounds, on average. While this is nearly twice the weight gain most frequently associated with smoking cessation, it assumes that only quitters gain weight when cigarette taxes increase. That is, it excludes the weight gain of those who merely reduce their cigarette consumption, but remain smokers.[34] Instead, suppose that one-half of the nearly 4,200 smokers in this quartile each gains five pounds. This would account for roughly three-quarters of the aggregate weight gain, implying that each recent quitter gained about six pounds. Moreover, if only half of recent quitters are induced to quit by higher taxes, the implied weight gain would be nearly thirteen pounds, which is more consistent with current estimates of the weight gain associated with smoking cessation. Finally, note that these are back-of-the-envelope calculations intended to show that the average gain seems reasonable, and not to make any claims about how much weight smokers gain when they are induced to quit smoking by higher taxes.

---

[34] It also excludes those who may gain weight independent of any change in smoking behavior due to negative income shocks, as discussed earlier.

VI. Conclusions

Using data from the National Health Interview Surveys, I find evidence that higher taxes led to weight gains for a set of likely female smokers, but not their male counterparts. Since weight gains may not harm the health of smokers, who are more likely than their non-smoking peers to be clinically underweight, I examine the distributional impacts of taxes and find evidence of gains in the right tail of the BMI distribution. In particular, I find increases in obesity among these women, but no effect on the fraction clinically underweight.

While this finding raises the possibility that, at least for women, higher cigarette taxes might, via increased levels of clinical obesity, reduce the aggregate health benefits of smoking cessation, it is highly likely that quitting remains a worthy goal for individual smokers. The purpose of this work is to investigate an unintended consequence that may or may not have real effects on health. Nevertheless, investigation of the possibility is relevant, especially in the context of the very large tax increases of the past two years, many of which roughly match the levels called for in the *Healthy People 2010* report. Ultimately, such variation should allow for additional and improved tests of this phenomenon.

Appendix A—Tax-induced income effects on body weight

A somewhat stylized example may be useful in summarizing the preceding discussion. Suppose that body weight is determined by the following simplified relationship which varies across individuals:

$$W = F(S, C(I)M(S)),$$

where W is weight, S is the amount smoked, I is disposable income, and C(I)M(S) is total calories which is composed of the amount of food or number of "meals" consumed, M(S), and the number of calories per meal or "food quality", C(I). If S and I depend on cigarette taxes (T), then the full effect of cigarette taxes on weight, which itself varies across individual smokers, is given by:

$$\frac{dW}{dT} = F_1 \frac{\partial S}{\partial T} + F_2 \frac{\partial M}{\partial S} \frac{\partial S}{\partial T} C(I) + F_2 \frac{\partial C}{\partial I} \frac{\partial I}{\partial T} M(S)$$

The total effect of cigarette taxes on weight is comprised of three "pieces." The first piece is the direct effect of tax-induced changes in smoking on weight. Demand theory and the physiological evidence cited suggest this term will be non-negative. It could, for example, represent the slowing metabolism linked to reductions in smoking. The second piece is the effect of tax-induced changes in smoking via the amount of food consumed, M(S). Available evidence suggests that the sign of this term will also be non-negative as the effect of taxes on smoking is non-positive, the effect of smoking on the amount of food consumed is likely negative and body weight rises with caloric intake. The last piece is the effect of tax-induced changes in disposable income via food quality, C(I). Since food quality is likely a normal good and, again, body weight rises with caloric intake, the overall sign of this piece depends on the impact of cigarette taxes on disposable income. Clearly, higher cigarette taxes will raise the disposable incomes of those who respond by quitting. However, since cigarettes are an addictive good, many more smokers will respond to higher taxes by merely reducing their consumption on the intensive margin or maintaining current consumption levels. More precisely, if smokers are inelastic in their price responsiveness, they will experience negative income shocks.

Table 1. Selected baseline characteristics, by whether MSA experienced a tax increase

	Experienced Tax Increase	Did Not Experience Tax Increase	Absolute Difference
Body mass index	26.231	26.232	0.001
"Near" underweight	0.074	0.075	0.001
Underweight	0.018	0.025	0.007
"Near" obese	0.107	0.108	0.001
Obese	0.186	0.192	0.006
Cigarette tax, in cents (1997)	42.5	37.0	5.5
Cigarette tax, in cents (2001)	73.7	37.0	36.7

Note: With the exception of the last row, figures are based on 1997 data. "Near" underweight and "near" obese are my constructs and are defined as BMI between 18.5 and 20.5 and BMI between 28 and 30, respectively. "Underweight" corresponds to the fraction with BMI no greater than 18.5 and "obese" corresponds to the fraction of individuals with BMI of at least 30.

Fig 1A. Kernel density--Log BMI--Males--1997

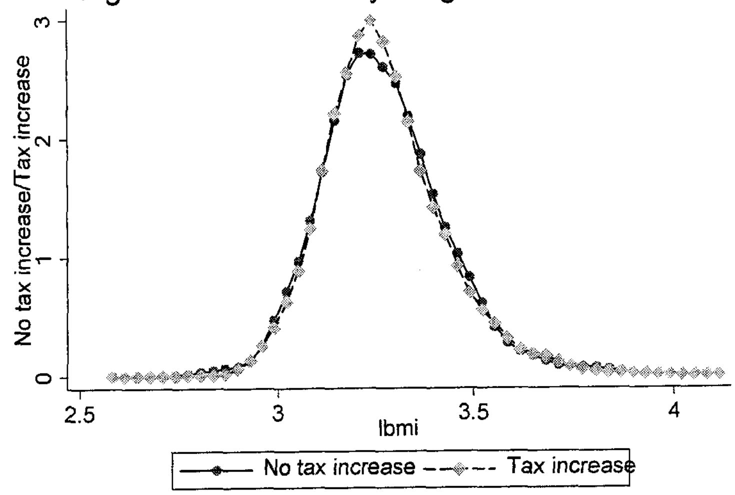

Fig 1B. Kernel density--Log BMI--Females--1997

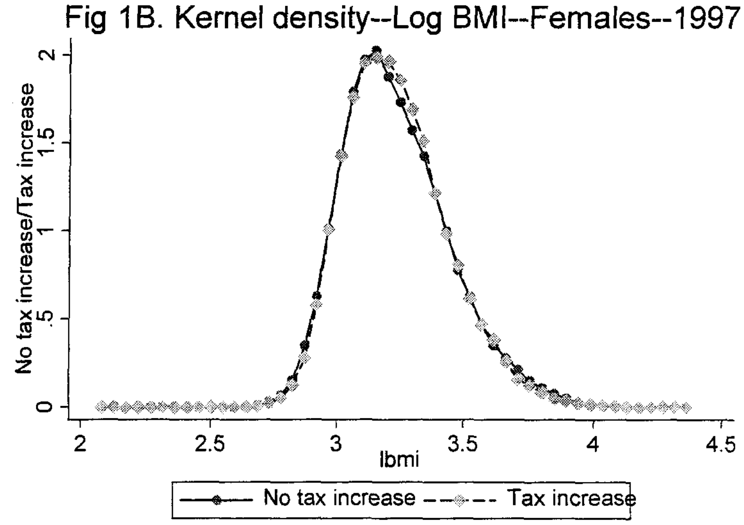

Figure 2--Divergence in tax regimes by MSA type, 1997-2001

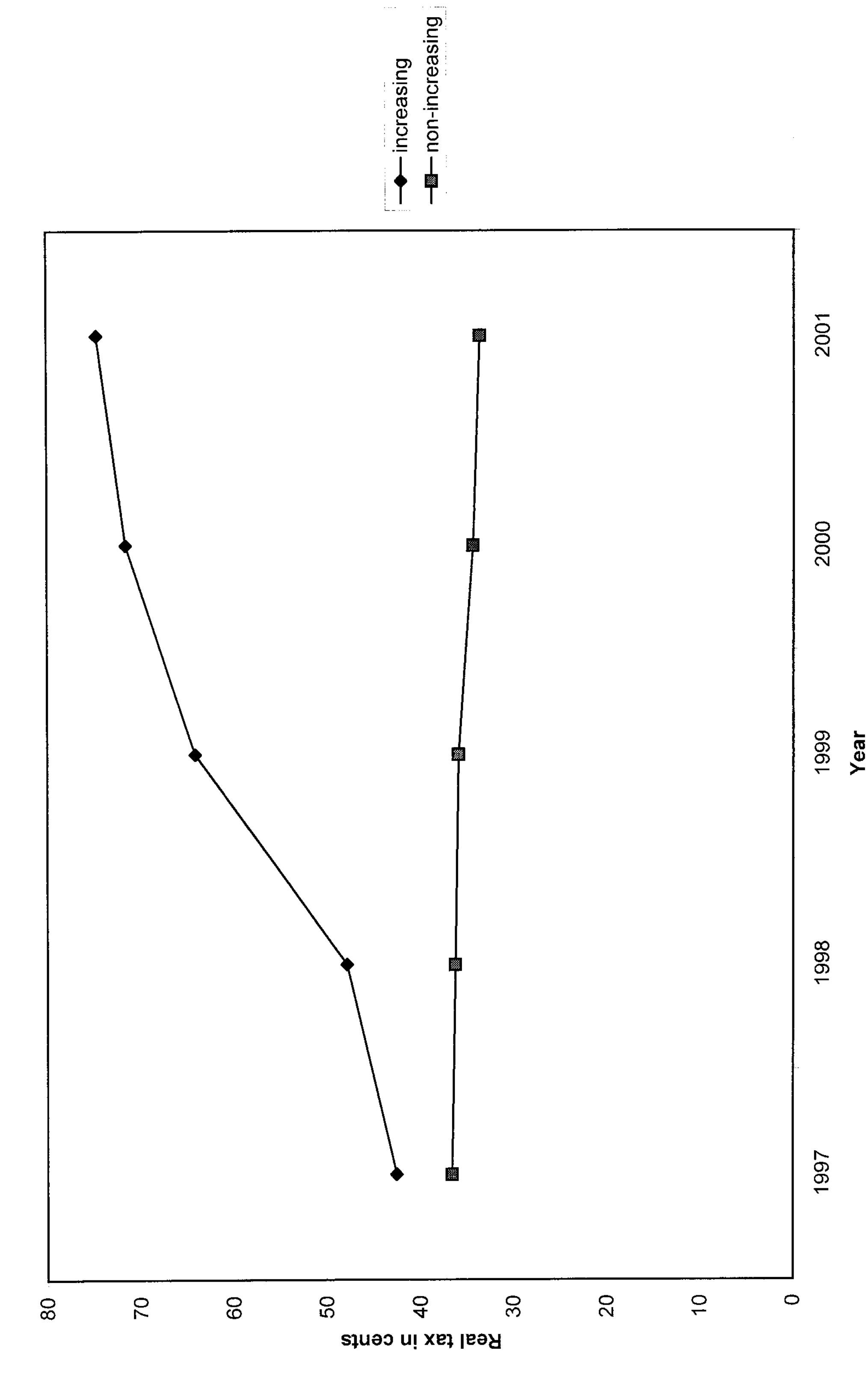

Figure 3A--Fraction male smokers by MSA type, 1997-2001

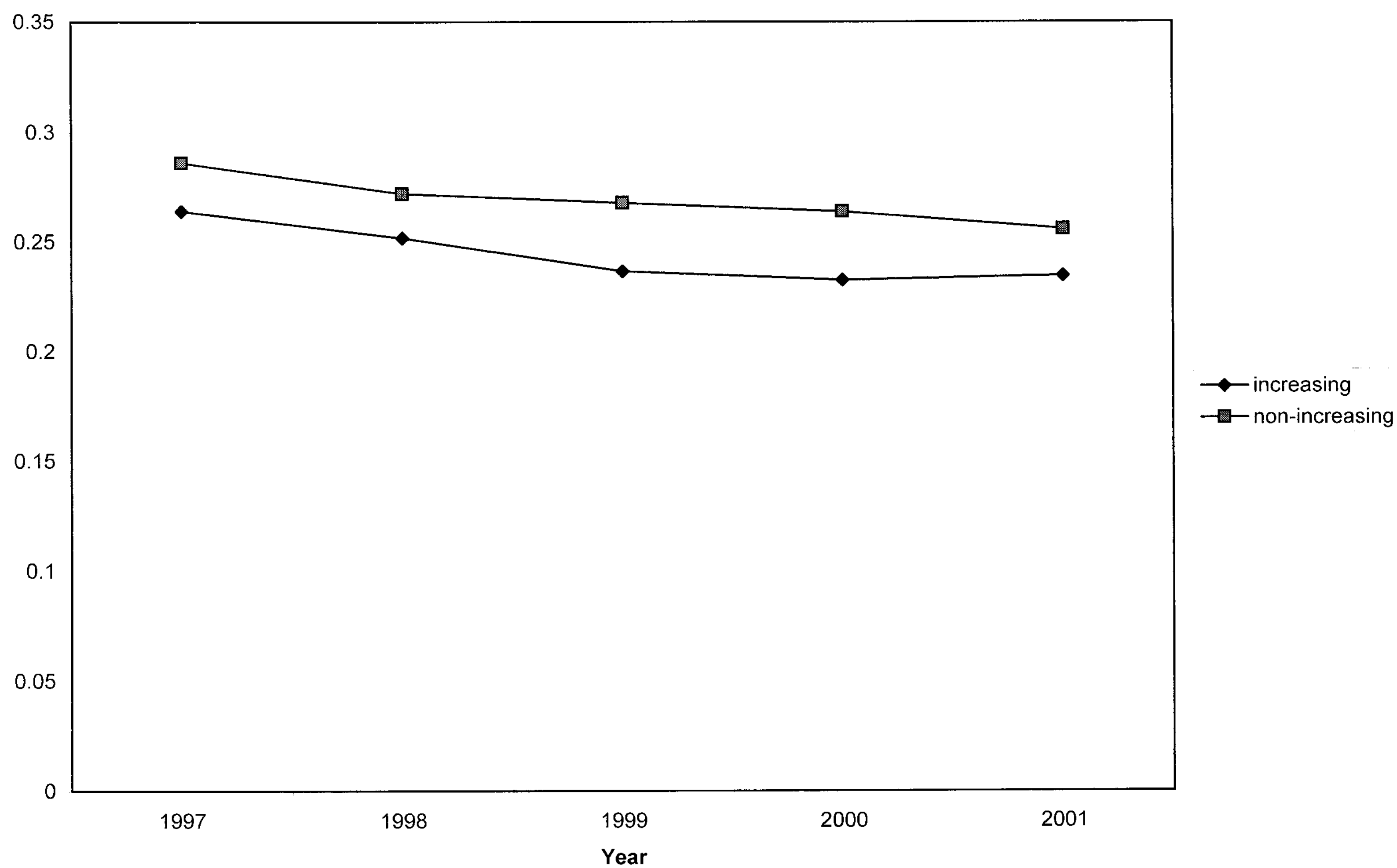

Figure 3B--Fraction female smokers by MSA type, 1997-2001

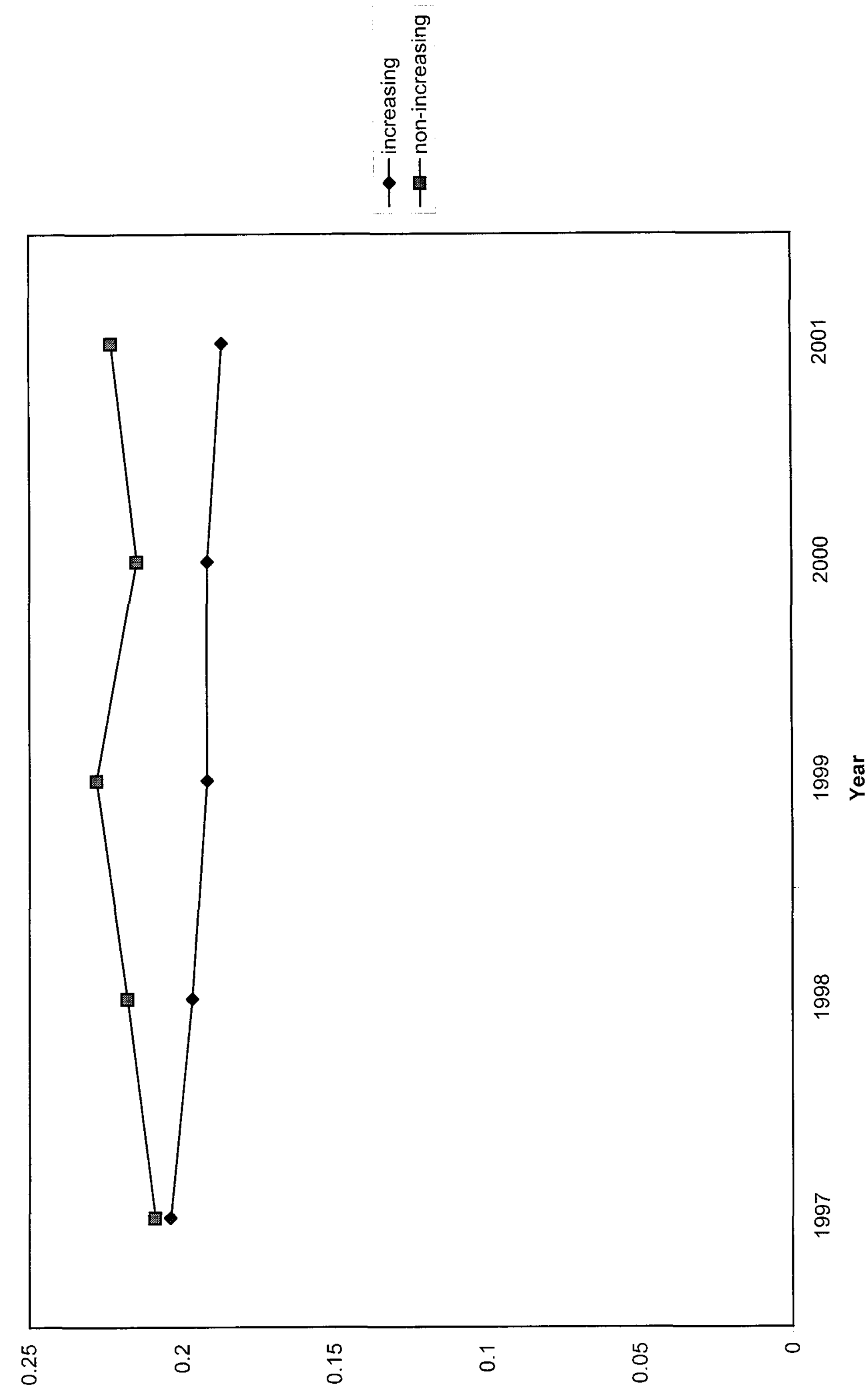

Table 2. Sample means, by quartile of predicted smoker distribution and gender

Quartile	1st	2nd	3rd	4th
Women				
Smoker	0.077	0.147	0.218	0.367
Heavy smoker	0.018	0.043	0.072	0.137
Cigarette tax	52.13	50.11	47.94	46.42
	(25.41)	(24.95)	(24.33)	(23.83)
BMI	24.94	25.74	26.49	27.21
	(5.09)	(5.54)	(6.20)	(6.57)
Obese	0.141	0.187	0.228	0.271
Underweight	0.040	0.031	0.028	0.028
Low income	0.074	0.100	0.168	0.343
Sample size	11,345	11,127	10,972	11,264
Men				
Smoker	0.103	0.193	0.289	0.437
Heavy Smoker	0.033	0.075	0.122	0.218
Cigarette tax	53.28	51.18	48.97	46.57
	(26.04)	(25.68)	(24.52)	(24.11)
BMI	26.33	26.94	26.94	26.98
	(4.11)	(4.57)	(4.85)	(4.99)
Obese	0.153	0.198	0.208	0.209
Underweight	0.008	0.006	0.009	0.009
Low income	0.043	0.080	0.127	0.199
Sample size	9,188	8,848	8,678	8,764

Notes: Heavy smokers report smoking at least one pack of cigarettes per day and low income refers to individuals with incomes below 125% of the federal poverty line. Sample sizes differ across quartiles because of missing BMI information in my analysis sample. Estimates generated from analysis sample which is described in data section of paper. Standard deviations in parentheses for continuous variables.

Table 3. Estimated effect of tax on log of BMI, by actual smoking status.

	Men	Women
Tax	-0.000042	0.000019
	(0.59)	(0.23)
Smoker	-0.048205	-0.054447
	(10.53)	(9.30)
Tax*Smoker	0.000161	0.000309
	(1.74)	(2.45)
Dependent mean	3.278	3.243
Sample size	35,478	44,708

Notes: Dependent variable in both models is log of body mass index (BMI). BMI is defined as an individual's weight in kilograms divided by height in meters squared. Both models include controls for education, income relative to poverty line, race, age, marital status, employment status, unemployment rate as well as year-specific quarter of interview and MSA fixed effects. Absolute values of t-ratios in parentheses. Standard errors adjusted for non-independence of observations within MSAs.

Table 4. Estimated effect of tax on log of BMI, by quartile of predicted smoking distribution and gender.

Quartile	Men	Women
1st	-0.000052 (0.51) [9,188] {3.259}	0.000127 (0.87) [11,345] {3.198}
2nd	-0.000232 (1.35) [8,848] {3.280}	0.000186 (1.31) [11,127] {3.227}
3rd	0.000130 (0.88) [8,678] {3.279}	0.000130 (0.81) [10,972] {3.253}
4th	0.000098 (0.62) [8,764] {3.279}	0.000317 (2.04) [11,264] {3.278}

Notes: Dependent variable is log of body mass index (BMI). BMI is defined as an individual's weight in kilograms divided by height in meters squared. Models include controls for education, income relative to poverty line, race, age, marital status, employment status, unemployment rate as well as year-specific quarter of interview and MSA fixed effects. Sample sizes, in square brackets, differ across quartiles because of differential missingness of BMI information by quartile. Absolute values of t-ratios in parentheses and average log BMI in curly brackets. Standard errors adjusted for non-independence of observations within MSAs.

Table 5A. Estimated effect of tax on selected BMI thresholds for men.

	1 (BMI≤16.5)	2 (BMI≤18.5)	3 (BMI≤20.5)	4 (BMI≥28)	5 (BMI≥30)	6 (BMI≥32)
Quartile						
1st	-0.000026	0.000130	-0.000144	-0.000074	-0.000164	-0.000300
	(1.00)	(1.55)	(1.01)	(0.21)	(0.67)	(1.51)
	{0.001}	{0.01}	{0.04}	{0.28}	{0.15}	{0.08}
2nd	0.000019	0.000068	0.000202	-0.000762	-0.000221	-0.000471
	(0.54)	(0.76)	(1.35)	(1.89)	(0.56)	(1.36)
	{0.002}	{0.01}	{0.04}	{0.34}	{0.20}	{0.12}
3rd	-0.000053	-0.000102	-0.000231	0.000796	-0.000213	-0.000126
	(1.45)	(1.37)	(1.34)	(1.68)	(0.57)	(0.51)
	{0.001}	{0.01}	{0.05}	{0.34}	{0.21}	{0.13}
4th	0.000024	0.000018	-0.000213	0.000116	0.000031	0.000159
	(0.79)	(0.24)	(1.22)	(0.26)	(0.09)	(0.52)
	{0.002}	{0.01}	{0.05}	{0.34}	{0.21}	{0.13}

Notes: Each column corresponds to a linear probability model where the dependent variable equals one if the BMI threshold indicated is met, and zero otherwise. Models include controls for education, income relative to poverty line, race, age, marital status, employment status, unemployment rate as well as year-specific quarter of interview and MSA fixed effects. Sample sizes for the quartiles are as follows: 9,188, 8,848, 8,678 and 8,764. They differ across quartiles due to missing body mass index information in my analysis sample. Absolute values of t-ratios in parentheses and proportion meeting the indicated threshold in curly brackets. Standard errors adjusted for non-independence of observations within MSAs.

Table 5B. Estimated effect of tax on selected BMI thresholds for women.

	1 (BMI≤16.5)	2 (BMI≤18.5)	3 (BMI≤20.5)	4 (BMI≥28)	5 (BMI≥30)	6 (BMI≥32)
Quartile						
1st	-0.000065	-0.000075	-0.000358	0.000057	0.000110	-0.000011
	(0.87)	(0.37)	(1.28)	(0.20)	(0.39)	(0.05)
	{0.01}	{0.04}	{0.17}	{0.22}	{0.14}	{0.09}
2nd	0.000015	-0.000024	-0.000152	0.000387	0.000223	-0.000075
	(0.32)	(0.21)	(0.50)	(1.10)	(0.73)	(0.30)
	{0.004}	{0.03}	{0.14}	{0.28}	{0.19}	{0.12}
3rd	0.000009	-0.000066	0.000150	0.000431	0.000126	0.000265
	(0.13)	(0.47)	(0.65)	(0.86)	(0.33)	(1.11)
	{0.004}	{0.03}	{0.12}	{0.33}	{0.23}	{0.16}
4th	0.000044	-0.000006	-0.000240	0.001218	0.000723	0.000599
	(1.06)	(0.05)	(0.87)	(3.26)	(2.01)	(2.17)
	{0.004}	{0.03}	{0.11}	{0.38}	{0.27}	{0.19}

Notes: Each column corresponds to a linear probability model where the dependent variable equals one if the BMI threshold indicated is met, and zero otherwise. Models include controls for education, income relative to poverty line, race, age, marital status, employment status, unemployment rate as well as year-specific quarter of interview and MSA fixed effects. Sample sizes for the quartiles are as follows: 11,345, 11,127, 10,972 and 11,264. They differ across quartiles due to missing body mass index information in my analysis sample. Absolute values of t-ratios in parentheses and proportion meeting the indicated threshold in curly brackets. Standard errors adjusted for non-independence of observations within MSAs.

Table 6. Fraction above desired weight, by predicted smoking quartile and gender, 1997

	Women			Men		
% above desired weight	5%	10%	20%	5%	10%	20%
Quartile						
1st	0.486	0.368	0.228	0.686	0.546	0.291
2nd	0.542	0.439	0.300	0.711	0.578	0.352
3rd	0.590	0.501	0.340	0.727	0.593	0.366
4th	0.647	0.561	0.418	0.714	0.592	0.367

Notes: These figures represent the proportion who report being five, ten or twenty percent above their desired weight. To avoid any effect of taxes on subjective weight, I limit the sample to 1997 data. Figures based on available data from members of analysis sample.

Table 7. Estimated effect of tax on smoking status, by gender

	Men	Women
1st Quartile	-0.000015 (0.06) {-0.058} [9,188]	0.000073 (0.38) {0.372} [11,345]
2nd Quartile	0.000116 (0.37) {0.231} [8,848]	-0.000139 (0.75) {-0.358} [11,127]
3rd Quartile	0.000022 (0.08) {0.028} [8,678]	0.000315 (1.26) {0.538} [10,972]
4th Quartile	0.000023 (0.07) {0.018} [8,764]	-0.000512 (1.57) {-0.486} [11,264]
All quartiles combined	0.000017 (0.11) {0.025} [35,478]	-0.000186 (1.29) {-0.340} [44,708]

Notes: Equations are estimated as linear probability models. Dependent variable is whether or not an individual is a smoker. Other covariates, not listed, include age, education, ratio of household income to poverty level, race, marital status, employment status, unemployment rate, as well as year-specific quarter of interview and MSA fixed effects. Absolute value of t-ratios in parentheses, sample sizes in square brackets and implied price participation elasticities in curly brackets. Standard errors adjusted for non-independence of observations within MSAs.

Table 8. Estimated effect of tax on the number of cigarettes smoked per day for current smokers, by gender.

	Men	Women
1st Quartile	0.027014	-0.006069
	(1.24)	(0.22)
	{0.861}	{-0.217}
2nd Quartile	-0.015804	-0.015770
	(0.89)	(0.88)
	{-0.425}	{-0.485}
3rd Quartile	-0.047562	0.003879
	(2.52)	(0.23)
	{-1.129}	{0.105}
4th Quartile	-0.030727	-0.001137
	(2.32)	(0.07)
	{-0.614}	{-0.027}
All quartiles combined	-0.018472	-0.006570
	(1.96)	(0.92)
	{-0.427}	{-0.180}

Notes: Sample consists of current smokers of whom there are 8,884 males and 9,070 females across the four quartiles. Dependent variable is average number of cigarettes smoked per day. Other covariates, not listed, include age, education, ratio of household income to poverty level, race, marital status, employment status, unemployment rate, as well as year-specific quarter of interview and MSA fixed effects. Absolute value of t-ratios in parentheses and implied price elasticities in curly brackets. Standard errors adjusted for non-independence of observations within MSAs.

Table 9. Estimated effect of tax on log of body mass index, underweight status and obesity status for highest quartile of predicted smoking distribution, by gender.

	1 (Log BMI)	2 (BMI≤18.5)	3 (BMI≥30)
Excluding three MSAs			
Men	0.000084	0.000030	-0.000031
(N=8,426)	(0.53)	(0.39)	(0.09)
	{3.279}	{0.01}	{0.21}
Women	0.000327	0.000005	0.000787
(N=10,697)	(2.08)	(0.04)	(2.21)
	{3.277}	{0.03}	{0.27}
Minimum tax strategy			
Men	0.000099	0.000046	0.000035
(N=8,764)	(0.66)	(0.63)	(0.10)
	{3.280}	{0.01}	{0.21}
Women	0.000401	-0.000036	0.000831
(N=11,264)	(2.60)	(0.27)	(2.40)
	{3.278}	{0.03}	{0.27}
Excluding three MSAs & Minimum tax strategy			
Men	0.000097	0.000049	0.000020
(N=8,426)	(0.66)	(0.65)	(0.05)
	{3.279}	{0.01}	{0.21}
Women	0.000398	-0.000037	0.000874
(N=10,697)	(2.50)	(0.27)	(2.48)
	{3.277}	{0.03}	{0.27}

Notes: The top panel corresponds to models that exclude the three most troublesome MSAs, the middle panel corresponds to models that assign the minimum state tax to the fourteen (of fifty-eight) MSAs with multiple state overlap and the bottom panel corresponds to models that impose both refinements. In each panel, the first column corresponds to an OLS regression where the dependent variable is log of body mass index. The second and third columns correspond to linear probability models where the dependent variable equals one if the BMI threshold indicated is met, and zero otherwise. Models include controls for education, income relative to poverty line, race, age, marital status, employment status, unemployment rate as well as year-specific quarter of interview and MSA fixed effects. Absolute values of t-ratios in parentheses and dependent means in curly brackets. Standard errors adjusted for non-independence of observations within MSAs.

References

Cawley, J.H., Markowitz, S. and J. Tauras (2004) "Lighting up and slimming down: The effects of body weight and cigarette prices on adolescent smoking initiation", *Journal of Health Economics*, 23: 293-311.

Cawley, J.H. (1999) "Rational addiction, the consumption of calories, and body weight", Ph.D. dissertation. University of Chicago.

Chaloupka, F.J. and K.E. Warner (2000) "The Economics of Smoking" In Culyer, A.J. and J.P. Newhouse, Eds. *Handbook of Health Economics*, vol. 1B. Amsterdam: Elsevier, 1539-1627.

Chou, S-Y., Grossman, M. and H. Saffer (2004) "An economic analysis of adult obesity: Results from the BRFSS", *Journal of Health Economics* 23(1): 565-587.

Clearman, D.R. and D.R. Jacobs (1991) "Relationships between weight and caloric intake of men who stop smoking", *Addictive Behaviors* 16: 401-410.

Cutler, D.M, Glaeser, E.L., and J.M. Shapiro (2003). "Why have Americans become more obese?", *Journal of Economic Perspectives* 17, 93-118.

Duan, N., Manning, W.G., Morris, C.N., and J.P. Newhouse (1983) "A comparison of alternative models for the demand for medical care", *Journal of Economic and Business Statistics*, 1: 115-126.

Evans, W.N. and J.S. Ringel (1999) "Can higher cigarette taxes improve birth outcomes?", *Journal of Public Economics* 72: 135-154.

Evans, W.N. and M.C. Farrelly (1998) "The compensating behavior of smokers: Taxes, tar and nicotine", *RAND Journal of Economics*, 29(3): 578-595.

French, S.A. and R.W. Jeffrey (1995) "Weight concerns and smoking: A literature review", *Annals of Behavioral Medicine* 17: 234-244.

Glauser, S.C. et al. (1970) "Metabolic changes associated with the cessation of cigarette smoking", *Archives of Environmental Health* 20: 377-381.

Gritz, E.R. et al. (1989) "The smoking and body weight relationship: Implications for intervention and post-cessation weight control", *Annals of Behavioral Medicine* 11: 144-153.

Colman, G, Grossman, M and T. Joyce (2003) "The effect of cigarette taxes on smoking before, during and after pregnancy", *Journal of Health Economics* 22:1053-1072.

Gruber, J. (2001) "Tobacco at the cross-roads: The past and future of smoking regulation in the United States", *Journal of Economic Perspectives* 15: 193-212.

Gruber, J. and S. Mullainathan (2002) Do cigarette taxes make smokers happier?, NBER working paper, #8872.

Gruber, J. and B. Koszegi (2001). "Is addiction rational? Theory and Evidence", *Quarterly Journal of Economics*, 116(4): 1261-1303.

Gruber, J. and B. Koszegi (2004) "Tax incidence when individuals are time-inconsistent: The case of cigarette excise taxes", *Journal of Public Economics*, 88: 1959-1988.

Grunberg, N.E. (1982) "The effects of nicotine and cigarette smoking on food consumption and taste preferences", *Addictive Behaviors* 7: 317-331.

Hall, S.M. et al. (1989) "Changes in food intake and activity after quitting smoking", *Journal of Consulting & Clinical Psychology* 57: 81-86.

Harris, J.E. (1987) "The 1983 increase in the federal cigarette excise tax", In Summers, L.H., editor, *Tax Policy and the Economy*, Vol. 1, Cambridge, MA: MIT Press, 87-111.

Harris, J.E. (1980) "Taxing tar and nicotine", *American Economic Review* 70(3): 300-311.

Hofstetter, A. et al. (1986) "Increased 24-hour energy expenditure in cigarette smokers", *New England Journal of Medicine* 314: 79-82.

Kershbaum et al. (1966) "Differences in effects of cigar and cigarette smoking on free fatty acid mobilization and catecholamine excretion", *Journal of the American Medical Association* 195: 1095-1098.

Klesges, R.C. et al. (1998) "The prospective relationships between smoking and weight in a young, biracial cohort", *Journal of Consulting & Clinical Psychology* 66: 987-993.

Klesges, R.C. et al. (1989) "Smoking, body weight, and their effects on smoking behavior: A comprehensive review of the literature", *Psychological Bulletin* 106: 204-230.

Klesges, R.C. and L.M. Klesges (1988) "Cigarette smoking as a dieting strategy in a university population", *International Journal of Eating Disorders*, 7(3): 413-419.

Lakdawalla, D and T. Philipson (2002) "The growth of obesity and technological change: A theoretical and empirical investigation", NBER Working Paper #8965.

Mokdad, A.H., Marks, J.S., Stroup, D.F. and J.L. Gerberding (2004) "Actual causes of death in the United States, 2000", *Journal of the American Medical Association* 291: 1238-1245.

Moore, M.J. (1996) "Death and tobacco taxes", *RAND Journal of Economics*, 27(2): 415-428

Ogden, J. and P. Fox (1994) "Examination of the use of smoking for weight control in restrained and unrestrained eaters", *International Journal of Eating Disorders* 16: 177-185.

Orzechewski, W. and R. Walker (2002) *The Tax Burden on Tobacco*, volume 35 Orzechewski and Walker, Arlington, VA.

Perkins, K.A. (1993) "Weight gain following smoking cessation", *Journal of Consulting & Clinical Psychology* 61: 768-777.

Perkins, K.A. et al. (1990) "Changes in energy balance following smoking cessation and resumption in women", *Journal of Consulting & Clinical Psychology* 58:121-125.

Pirie, P.L., Murray, D.M. and R.V. Luepker (1991) "Gender differences in cigarette smoking and quitting in a cohort of young adults", *American Journal of Public Health*, 81(3): 324-327.

Philipson, T. and R.A. Posner (1999) "The long-run growth in obesity as a function of technological change", NBER Working Paper #7423.

Pomerleau, O.F. et al. (1991) "Effects of fluoxetine upon weight gain and food intake in smokers who reduce nicotine intake", *Psychoeuroendocrinology* 16: 433-440.

Rodin, J. (1987) "Weight change following smoking cessation: The role of food intake and exercise", *Addictive Behaviors* 12:303-317.

Seidell et al. (1996) "Overweight, underweight and mortality: A prospective study of 48,287 men and women", *Archives of Internal Medicine,* 156: 958-963.

Stamford, B.A. et al. (1986) "Effects of smoking cessation on weight gain, metabolic rate, caloric consumption and blood lipids", *American Journal of Clinical Nutrition* 43: 486-494.

U.S. Department of Health and Human Services [USDHHS] (1990) *The health benefits of smoking cessation: A report of the Surgeon General* (DHHS Publication No. 90-8416). Washington, DC: U.S. Government Printing Office.

U.S. Department of Health and Human Services [USDHHS] (2000) *Reducing tobacco use: A report of the Surgeon General.* Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office of Smoking and Health.

U.S. Department of Health and Human Services [USDHHS] (2004) *Data 2010: The Healthy People 2010 Database*. Division of Health Promotion Statistics, National Center for Health Statistics.

Wack, J.T. and J. Rodin (1982) "Smoking and its effects on body weight and systems of caloric regulation", *American Journal of Clinical Nutrition* 35: 366-380.

Warner, K.E. (1986) "Smoking and health implications of a change in the federal cigarette excise tax", *Journal of the American Medical Association* 255(8): 1028-1032.

Williamson, D.F., et al. (1991) "Smoking cessation and severity of weight gain in a national cohort", *New England Journal of Medicine* 324: 739-745.

CHAPTER 3

DOES FULL-DAY KINDERGARTEN MATTER? EVIDENCE FROM THE FIRST TWO YEARS OF SCHOOLING

I. Introduction

Over the past three decades, the fraction of U.S. kindergartners enrolled in full-day programs has risen from roughly one-tenth to a slight majority (U.S. Census, 2002). Despite the extra cost of providing full-day kindergarten, it remains popular where it exists and is growing in popularity where it does not. To date, there exists little systematic evidence regarding its possible effects on academic achievement and even less information on their persistence over time. Using longitudinal data, I examine the impact of full-day kindergarten attendance on standardized test scores in mathematics and reading, as children progress from kindergarten to first grade.

I find evidence that full-day kindergarten has positive and practically important effects on early human capital formation. However, the estimated gains to full-day kindergarten are short-lived, as they fall dramatically over the course of an additional year, towards the end of first grade.[1] This pattern is especially striking for black and Hispanic full-day kindergartners who see significant short-run gains depreciate more completely than their white peers. Hispanic children, in particular, exemplify this pattern. I also find, among white children, larger and more persistent effects for plausibly

[1] These gains are based on measurements taken one year later, near the end of first grade. In what follows, "short-run" refers to *end-of-kindergarten* test scores and "longer-run" refers to *end-of-first grade* scores.

disadvantaged children. I find no such effects, however, for similarly disadvantaged minority children.

In the following section, I provide background on the question of interest, briefly discussing the history of full-day kindergarten and reviewing the relevant literature. I also briefly discuss why full-day kindergarten might affect children's human capital formation and provide some sample comparisons between full and half-day kindergartners in order to understand how much they differ on a detailed set of observed characteristics. As will be seen, full and half-day kindergartners appear strikingly similar today, due principally to the historical evolution of full-day kindergarten. Section III discusses the data, emphasizing their particular advantages in producing meaningful estimates of the relationship of interest. The basic empirical strategy, which exploits the fact that kindergartners were given mathematics and reading tests at multiple points in time, is also presented in this section. Most importantly, tests were first administered to children shortly after entering kindergarten, prior to much exposure to the treatment of interest. Throughout the paper, I refer to child performance on these initial tests as *baseline* scores. In conjunction with companion tests given toward the ends of kindergarten and first grade, these tests are designed to allow assessment of child learning in reading and mathematics. Section IV presents results from models estimated separately by race and also by race and gender in later specifications. I also briefly investigate the counterfactual. That is, I consider the possibility that the activities half-day children engage in when not in school are the reason for the estimated differences between full and half-day kindergartners, and not the extra schooling inherent in full-day kindergarten, itself. Section V provides two important extensions. First, I present evidence on the distributional impacts of full-day kindergarten. That is, do the gains associated with full-day kindergarten, if any, accrue mainly to disadvantaged children or do they serve to increase existing achievement gaps? Next, I provide evidence on *why* short-run effects disappear by focusing on *when* they disappear. Section VI concludes the paper.

II. Background and Motivation

A. The Movement (Back) to Full-Day Kindergarten

In the United States, the history of kindergarten dates back to the late 1800s, when it began as a full-day program. Kindergarten grew in popularity through the early twentieth century and remained full-day until the U.S. entered World War II in the 1940s. During the war, the need for labor in war-time industries drew many women into the labor force—some for the first time and some from other employment sectors (Goldin, 1991). Among the latter category were teachers, who at that time were almost uniformly women, especially in the elementary grades. As part of the larger "war effort", schools across the country began to cut their kindergarten classes back to half-day in order to free up additional labor (Oelerich, 1984; Jones, 2002). After WWII, tremendous growth in the number of young children (i.e., the early Baby Boom cohorts) reinforced the trend towards half-day kindergarten (Ulrey et al, 1982; Jones, 2002).

Kindergarten retained its half-day character until the 1960s and 1970s when full-day kindergarten began to reemerge as a way of improving the academic preparation of children deemed "at-risk".[2] As such, the early re-adopters of full-day kindergarten tended to be poorer schools and school districts, serving predominantly minority children. The prevailing view was that full-day kindergarten, which typically consists of a five to

[2] One notable exception is Hawaii, which reestablished full-day kindergarten in the 1950s (Gorton, 1968).

six hour school day, rather than the typical two and one-half to three hour day in half-day kindergarten, would provide more opportunity for these children to "catch up" academically to their less disadvantaged peers. Over time, many non-poor districts, perhaps in partial response to the rise in dual-earner and single-parent families, began to implement full-day programs. Today, full-day kindergarten is the norm, albeit by a slight margin. As I discuss in great detail later in this section, the variegated reemergence of full-day kindergarten has led to full and half-day kindergartners who are presently very similar on many dimensions.

Much of the very recent push for full-day kindergarten has occurred at the state level. Presently, nine state governments mandate full-day kindergarten and twenty-six provide financial incentives to encourage school districts to provide it. Much of the enabling legislation occurred in the 1990s. For example, of the nine mandating states, seven initiated full-day kindergarten-only regimes after 1990. Other states may soon follow as evidenced by the comments of Nevada Governor Kenny Guinn in his 2003 State of the State address:

> "...to realize that vision we need to create a generation of young Nevadans with stronger, sharper, and more sophisticated skills. Therefore, I propose that we start at the beginning by providing full-day kindergarten for our children."

Full-day kindergarten remains popular, despite being much more expensive than the half-day variety. These higher costs are somewhat intuitive. Consider a school that is required to switch from traditional half-day to full-day kindergarten. Holding constant class size, the number of days per week that class meets, the length of the school year and other potentially offsetting (i.e., cost-saving) behavior, this school now requires twice as many teachers for a fixed number of pupils. In addition, full-day kindergarten offers fewer opportunities to share resources such as desks, books and computers relative to a half-day regime. Finally, full-day kindergarten may involve higher quasi-fixed costs. For example, a school might need additional classroom space since it will no longer be able to use the same room two times per day, as with a half-day regime. More concretely, Ohio's Office of Education Oversight (1997) estimates that full-day kindergarten costs over seventy percent more than traditional half-day kindergarten in terms of per pupil expenditure. The report notes that full-day kindergarten is not twice as expensive as half-day kindergarten primarily due to savings in transportation costs.[3]

B. Why Might Full-Day Kindergarten Matter?

The rationale for full-day kindergarten is simple: the more time children spend in school the more they will learn. In turn, it is thought that this additional learning will lead to improved academic outcomes as children move into later grades. In the context of a human capital accumulation story, early investments reduce the cost of future ones, so a larger initial or early stock of human capital has the potential to influence later, perhaps even adult, levels of human capital. More generally, if one fails to learn the "basics", it may inhibit subsequent learning so that "catching up" is prohibitively costly. As noted by Nobel laureate Gary S. Becker in his 1989 Ryerson Lecture:

> "Large differences among young children grow over time with age and schooling because children learn more easily when they are better prepared. Therefore, even small differences among children...are frequently multiplied over time into large differences when they are teenagers."

In the present case, children who leave kindergarten with relatively better reading skills are likely to be more successful in learning new material in the first grade and beyond, especially since the material taught in early elementary school tends to be sequential in nature (Siefert, 1993). Mathematics provides perhaps an even better

[3] In particular, schools must only bus children two times per day under a full-day regime, instead of four times per day when operating morning and afternoon sessions under half-day kindergarten.

example. Mathematical learning tends to be quite sequential in nature, so if one masters the basic concepts early it is likely that the burden of future learning will be lowered. To the extent that learning at this level is indeed sequential, it is possible that full-day kindergarten has effects that persist over time. In this spirit, recent work links the introduction of kindergartens in the South to increased educational attainment for blacks (Cascio, 2003). While not directly related to the present question, it suggests that additional early childhood education may have persistent impacts since final educational attainment is realized much later, typically in late adolescence or early adulthood. As described in section III, the longitudinal nature of my data allows me to test for persistence, albeit over a much shorter amount of time.

C. Related Literature

The preceding logic implicitly assumes that extra "seat time" provided by full-day kindergarten is devoted to learning activities. Available evidence lends support to this assumption. For example, Hough and Bryde (1996) and Elicker and Mathur (1997) show that teachers in full-day kindergarten settings spend more time with children individually and in small groups, relative to teachers in half-day programs. As a result, full-day kindergartens are more able to integrate the on-going "push down" of academic material traditionally presented in the first grade since they have more time to incorporate these concepts and/or related learning activities (Elicker and Mathur, 1997). Full-day kindergarten, however, is not without its detractors. For example, some warn that an early emphasis on academic learning, at the expense of the traditional play-based curriculum, may harm children emotionally and, consequently, academically as well (Olsen and Zigler, 1989; Gullo, 1990; Natale, 2001).

While there has been considerable attention in the non-economics education literature, economists have shown little interest in the full vs. half-day kindergarten debate. The existing literature suggests that full-day kindergarten's impact on academic and social outcomes is somewhat mixed, but taken as a whole tends to imply that full-day kindergarten's pros outweigh its cons. Some studies find relatively large gains (Gullo, 2000; Fusaro, 1997; Cryan et al, 1992), while others do not (Karweit, 1992; Puelo, 1988). It should be noted, however, that existing work has several limitations. For example, several studies focus on the experiences of a particular school or school district, some focus on particular types of students and others ignore concerns related to non-random sorting.

Given the abundance of closely related work, the lack of attention by economists is surprising. In particular, much attention has been given to the pre-kindergarten program, Head Start, and its possible effects on child academic, social and health outcomes. As recently as the middle and late 1990s, economists studying Head Start have linked participation with higher test scores, reduced grade repetition and an increased likelihood of receiving recommended childhood immunizations (Currie and Thomas, 1995; Currie and Thomas, 1999). Even more recently, Garces et al. (2002) find evidence that Head Start effects persist into early adulthood. Many of the reasons that one might suspect Head Start to improve child outcomes also apply to full-day kindergarten.[4]

[4] Note, however, that Head Start programs often involve non-academic interventions (e.g., promotion of available social services or parenting skills classes), while full-day kindergarten focuses more narrowly on classroom learning. As a result, I consider only learning-related outcomes, unlike the Head Start literature discussed above. Note also that all models with covariates include controls for prior Head Start participation.

D. Comparing Full and Half-Day Kindergartners: How Much Do They Differ?

A common concern in a study such as this is that "treated" and "untreated" individuals differ on unobserved traits that are correlated with treatment status and also independently affect the outcome of interest. To the extent that this is true, traditional estimators may yield misleading estimates of program effects. In the present context, the concern is that children who attend full-day kindergarten differ appreciably, and unobservably, from half-day children. And while it is impossible to compare children on the basis of unobserved characteristics, some observable comparisons may be suggestive of the degree of non-random sorting into kindergarten type. At a minimum, one can ascertain if the two groups are so different as to make comparison very difficult or perhaps even meaningless.

To reduce unobserved heterogeneity, I restrict my sample to children enrolled in public schools, who are first-time kindergartners and who did not change schools during kindergarten. I also perform the analyses by race group to further reduce this sort of heterogeneity (e.g., unobserved school quality). Table 1 provides several comparisons by kindergarten type for white, black and Hispanic children, respectively.[5] The first two columns of Table 1 show that among white children, the two types of kindergartners are strikingly similar. Perhaps most importantly, mean baseline mathematics and reading scores, as well as their corresponding standard deviations, are virtually identical and not statistically different from each other. In particular, the mean half-day math score is 1.2 percent higher than the full-day average and the corresponding reading score difference is less than 0.05 percent, with respective absolute t-ratios of 1.26 and 0.05. Since tests are

[5] The samples used to produce the figures in Table 1 correspond exactly to the samples upon which model estimates in Tables 2 through 4, which I present in section IV, are generated.

designed to yield normally distributed scores, the sample moments in Table 1 suggest that white full and half-day kindergartners belong to very similar baseline test score distributions. Beyond baseline scores, the fraction of children residing with two parents, the fraction whose families ever participated in AFDC/TANF and the fraction of kindergartners with working mothers are all quite similar. There are more substantial differences with respect to family income and parents' education, though these suggest a degree of "negative" selection into full-day kindergarten, but the differences are, practically speaking, small.[6]

Among black and Hispanic children, the means suggest that the two groups are slightly less similar than full and half-day whites. As seen in the middle two columns of Table 1, black full-day kindergartners enjoy a three percent advantage in the baseline reading test score, but have an average baseline mathematics score which is 0.4 percent lower than their half-day peers. Neither difference is statistically significant at conventional levels ($t=0.17$ and $t=1.42$, respectively). While the reading-score advantage suggests some degree of positive selection for black full-day kindergartners, the education level and income of the parents of black full-day kindergartners imply a slight degree of negative selection. With respect to baseline test scores, the degree of positive selection into full-day kindergarten is greatest among Hispanic children. The last two columns of Table 1 show that Hispanic full-day kindergarteners have baseline reading scores that are four percent higher than their half-day peers. However, corresponding mathematics scores are only 1.9 percent higher. Once again, neither difference in means

[6] Also, while there is little difference in maternal employment between birth and kindergarten, mothers of white children in full-day kindergarten are more likely to work thirty-five or more hours per week. However, the direction of causality may be from kindergarten type to full vs. part-time work (c.f., Lemke et al., 2000). Much smaller employment differences are seen for the mothers of minority children. All models estimated include a series of indicators for mother's employment status.

is statistically different from zero at conventional levels, though the reading score difference could be considered marginally significant (t=0.92 and t=1.75 for math and reading, respectively). As with black children, there is also some evidence that full-day kindergartners of Hispanic origin are negatively sorted into this status. So, while minority full and half-day kindergartners are not as similar as white full and half-day children, the differences are relatively minor.

III. Data and Empirical Strategy

A. Key Variables

The variables of greatest interest are the math and reading test scores, which are my outcomes of interest, and kindergarten status. As noted earlier, these tests were administered to children at multiple points in time. Of greatest importance to this study is that tests were given early in the first year of schooling, before much exposure to the kindergarten curriculum. Following these baseline tests, students were then re-assessed at two later points in time—towards the end of the kindergarten year and towards the end of first grade. This testing structure was designed explicitly to measure children's longitudinal gains in subject-specific achievement (NCES, 2002b). Since the amount of time between tests may influence achievement gains, all models include separate month of assessment indicators.

Though the tests differ over time, they contain common and overlapping elements. The mathematics tests assess number recognition, counting, comparing and ordering numbers, solving word problems and interpreting simple graphs. The reading tests include questions to assess the basic reading skills, vocabulary/word comprehension, knowledge of the alphabet, phonetics, listening skill and reading comprehension.[7] While a variety of scores based on these tests are available in the ECLS-K, the analysis presented here uses Item Response Theory (IRT)-adjusted scores rather than, for example, the raw number of correct answers provided. These particular scores adjust for the fact that the tests were not standardized per se, but instead asked different questions to different children, depending on their answers to a set of initial "routing" questions. This sort of adaptive testing is considered by psychometricians to be more efficient compared to pure standardized testing, where all students take the same examination, and also reduces the potential for "ceiling" and "floor" effects which can affect the measurements of gains over time (Lord, 1980). See the *ECLS-K Psychometric Report* and/or Chapter 3 of the *ECLS-K User's Guide*, especially pages 3-2 through 3-5, for more detailed information on these tests including test validity, reliability, differential item response and test-taker motivation (NCES, 2002b; NCES, 2002c).

Kindergarten type is the covariate of greatest interest. According to national data, nearly sixty percent of all kindergarteners were enrolled in full-day kindergarten in 2000 (U.S. Census, 2002). In the ECLS-K, which sampled children who entered kindergarten in academic year 1998-99, the corresponding fraction is approximately fifty-three percent. This slight difference may arise from the fact that the Census report, based on data from the October 2000 Current Population Survey, includes private schools while my sample includes only public school students. Indeed, the fraction enrolled in full-day kindergarten rises to fifty-seven percent when I include private school children.

---

[7] The tests were developed especially for the ECLS-K, but are based largely on existing and generally accepted instruments including the Children's Cognitive Battery (CCB), Peabody Individual Achievement Test—Revised (PIAT-R), Peabody Picture Vocabulary Test-3 (PPVT-3), Primary Test of Cognitive Skills (PTCS) and the Woodcock-Johnson Psycho-Educational Battery—Revised (WJ-R).

At the school-level, ninety-two percent of public schools (656 out of 714) offer either full or half-day kindergarten. Of these 714 schools, 348 (forty-nine percent) provide full-day kindergarten and 308 (forty-three percent) provide half-day kindergarten. The remaining fifty-eight schools (eight percent) provide both types. This preponderance of one type or the other, as opposed to offering both kindergarten types, is consistent with what can be gleaned from the literature on full-day kindergarten. More importantly, it suggests that within-school sorting is not widespread. That is, to the extent that kindergarten type is determined at the school-level and there is not as much scope for choosing full vs. half-day kindergarten within a particular school, there are fewer possibilities for non-random sorting since families would have to change (public) schools to change kindergarten type. Later, as a robustness check, I estimate models that exclude children in schools with both types of kindergarten and find very little difference relative to models that include children in all three school types. Finally, note that the "all or nothing" nature of full-day kindergarten effectively precludes the inclusion of school fixed effects in my models.

B. Empirical Strategy

The ECLS-K was designed, in part, to assess the value-added of kindergarten (NCES, 2002b). As such, standardized tests in math and reading were administered to nearly all sample members near the beginning and end of kindergarten. The existence of test scores prior to much exposure to schooling provides a baseline of the child's ability in these subject areas. Subsequent test scores, measured near the ends of kindergarten, first, third and fifth grades, provide the opportunity to assess both short and longer-run impacts of full-day kindergarten. I examine end-of-kindergarten and end-of-first grade scores in this paper.

Given the longitudinal nature of this information, the preferred statistical approach is one that accounts for individual heterogeneity. To fix ideas, consider test scores as an outcome of interest. If test score is regressed on kindergarten type in a standard cross-sectional model, a reasonable concern is omitted variable bias resulting from non-random sorting. For example, perhaps full-day children possess greater "readiness to learn" than their half-day counterparts *prior to* entering kindergarten, but this fact is not captured in the model. If so, the estimated impact of full-day kindergarten is likely to be overstated. However, if the impact of this unobserved difference is relatively constant over time, its effects should be embedded in the baseline test scores included in ECLS-K data. Hence, examining test score "growth" over time should eliminate, or at least mitigate, any associated bias.

I take a conceptually similar approach, controlling for the appropriate baseline test score as a right-hand side variable. More precisely, I estimate variants of the following general specification:

$$\mathbf{TS}_{i,t+j} = \psi\mathbf{TS}_{it} + \gamma\mathbf{FDK}_{it} + \alpha\mathbf{C}_{it} + \delta\mathbf{F}_{it} + \theta\mathbf{S}_{it} + \varepsilon_{i,t+j}, \qquad \mathbf{(1)}$$

where **TS** is subject-specific test score, **FDK** is a full-day kindergarten indicator, **C** is a set of child-specific variables, **F** is a set of family-specific variables and **S** is a set of school and classroom-specific variables and "j" determines whether the model estimates short or longer-run effects.[8] All models are estimated via ordinary least squares.

[8] To be clear, j=1 refers to end of kindergarten and j=2 refers to end of first grade.

Empirically, my goal is to estimate γ in the presence of this baseline score.[9] In essence, this is a more flexible way of estimating test score growth. Indeed, if ψ equals one, the model specified above would be equivalent to differencing the dependent variable, which would correspond to a difference-in-differences approach. The similarity of the two approaches is seen in Table A1, which compares estimates from a difference-in-differences model (without covariates) to my equation (1) with the restriction that $\boldsymbol{\alpha}=\boldsymbol{\delta}=\boldsymbol{\theta}=0$, so that the only covariates are baseline test score (TS) and kindergarten type (FDK). Table A1 reveals that estimates and their corresponding standard errors are quite similar across the two models, and are nearly identical in many cases. As will be seen in section IV, estimates in Table A1 also provide a fairly accurate preview of eventual results, especially their temporal pattern.

A couple of related items deserve mention. First, in practice, estimates of ψ are always relatively close to one. However, since ψ's are estimated very precisely, they are often statistically different from one, indicating that differencing the dependent variable may not be appropriate. Hence, I retain the more flexible specification implied by equation (1). Second, in all models presented, covariates are measured as of the initial survey wave out of necessity. While I observe test scores for the entire sample at three different points in time, I can not time difference covariates since only a small subset are measured in all three relevant waves of the ECLS-K. While certainly a methodological weakness, the relatively short period of time between the beginning of kindergarten and the end of first grade (roughly eighteen to twenty months), and the even shorter period

[9] I also estimate models with higher-order terms in baseline test score to allow for a more flexible functional form than is implied by equation (1) which imposes a linear relationship between baseline and subsequent test scores. The inclusion of these higher order terms had no appreciable impact on the relevant estimate of γ.

between the beginning and end of kindergarten (roughly nine months), imply that such differencing would result primarily in noisy measures of intertemporal change.

Finally, in addition to kindergarten type, several other potentially relevant school-level characteristics are available in the ECLS-K. I use information on school size, class size, length of the school year, number of days per week school is in session, public school type (i.e., regular, school of choice or magnet school) and whether the child's parents chose their current residence on the basis of the local schools.[10] To the extent that these school-level features are correlated with the type of kindergarten offered, and exert an independent effect on achievement, they represent important controls. Further, schools, especially those *required* to provide more expensive full-day kindergarten, may offset additional costs by increasing class size, reducing the length of the school year or the number of days class meets per week, etc., if they are able to do so.[11] However, to the extent that such offsetting behavior is directly attributable to full-day kindergarten, ignoring such school responses involves estimating a more general effect. In other words, the two estimates are conceptually different. Therefore, as noted, I estimate models with and without school-level covariates. In all cases, robust standard errors are adjusted for clustering at this level of aggregation (Moulton, 1990).

C. Analysis Sample

[10] This last variable is not measured at the school level, but I include it here since it is school-related.
[11] In my analysis sample, which I describe in great detail in the next sub-section, I find only small differences in class size (twenty vs. twenty-one students per class) and length of school year (177 vs. 178 days per year) for full and half-day kindergartners, respectively. With respect to days of class per week, however, there is some evidence of offsetting behavior. In particular, while only one percent of half-day children attend school less than five days per week, roughly six percent of full-day children do so.

While four waves of the ECLS-K were gathered by the end of first grade, I use only three—fall kindergarten (1998), spring kindergarten (1999) and spring first grade (2000)—since the fourth, fall first grade (1999), deliberately sampled only thirty percent of original respondents. Restricting my sample to those with math and reading scores in all three waves, discernible kindergarten type, and a completed initial parent questionnaire yields a sample of 13,025 children. Further restricting the sample to public school children who are first-time kindergarteners and do not switch schools during the kindergarten year leaves a sample of 9,632 children. Since analysis is done separately by race, I limit the sample to those races with enough sample size to support estimation by gender. Doing so produces an analysis sample of 8,599 children, of whom 5,785 are white, 1,583 are black and 1,231 are Hispanic.

As previously mentioned, I estimate models with and without school-level covariates. Without the school-level characteristics, complete case analysis yields a sample of 7,303, which includes 5,075 white, 1,216 black and 1,012 Hispanic children. With the school-level covariates included, the corresponding sample size drops to 5,734, including 4,189 white, 861 black and 684 Hispanic children. Most of the reduction in sample size is due to missing information on household income and school characteristics.[12] To understand the potential sensitivity of my results to the impact of missing data, I include separate binary indicators for those individuals with missing information on these covariates.[13] This increases sample size to 8,164, which includes 5,559 white, 1,445 black and 1,160 Hispanic children. These numbers represent 96

[12] Of the analysis sample (8,599 children), about eleven percent are missing household income information and about twelve percent are missing various school characteristics. No other covariate is missing for more than three percent of cases.
[13] As part of this strategy, I converted the variables household income, length of school year, number of days per week school meets and class size each into a series of discrete indicators.

percent (5,559 of 5,785), 91 percent (1,445 of 1,583) and 94 percent (1,160 of 1,231) of the white, black and Hispanic analysis samples, respectively. Since estimates, and their pattern, are very similar across the different samples, I report results from models that use this latter, and most complete, sample. [14]

IV. Results

In what follows, I refer frequently to short and longer-run estimates of the impact of full-day kindergarten. Recall that *short-run* refers to performance on an end-of-kindergarten reading or math test and that *longer-run* refers to performance on a similar test, designed to evaluate academic progress, roughly one year later. Since I estimate models separately by race group, I report estimates in a similar fashion. In each race-specific sub-section, I present short-run results followed by corresponding longer-run estimates. Further, since the literature on child cognitive development suggests the possibility of differences in the learning patterns of girls and boys, I also present estimates by gender for each race group (c.f., Fennema and Sherman, 1977; Carr and Jessup, 1997).

Tables 2 through 4 present regression estimates from three different specifications—Column 1 presents results from a simple regression of test score on kindergarten type, Column 2 presents a regression that adds an extensive set of covariates, but not the subject-specific baseline test score, and Column 3 presents a model that adds the appropriate baseline test score to the model in Column 2.[15] All

[14] After reading the results presented in the next section, please refer to Tables A2A and A2B for evidence regarding the robustness of full-day kindergarten estimates across different samples and model specifications. See corresponding table notes for a detailed explanation of what is presented.
[15] This set up allows the reader to compare the estimated effect of full-day kindergarten via a simple conditional mean difference, a naïve model that does not control for the appropriate baseline test score, and finally a model that also includes the baseline score. While my data are not experimental by design, estimates of the effect of full-day kindergarten are very close across these three specifications, especially

specifications are variants of equation (1). In all cases, Column 3 contains my preferred specification. These three tables each include two tables (e.g., Table 2A and Table 2B), where the former displays short-run estimates and the latter presents longer-run estimates. Both sets of estimates are generated from balanced samples.[16] Table 5 presents short and longer-run estimates by gender, but only for the preferred specification.

Given the large volume of estimates, I limit discussion to the estimated impacts of full-day kindergarten, sometimes to the exclusion of other potentially interesting results. In all cases, estimated differences in the performance of full-day kindergartners relative to their half-day counterparts are presented as a percentage of the standard deviation of the model-specific dependent variable and are discussed in terms of "gains" or "advantages".[17] Discussion centers on estimates from my most preferred specification. Finally, since there is little difference in the estimated impact of full-day kindergarten with and without them, all models I discuss include school-level covariates.

A. White Children

Short-run estimates, contained in Table 2A, imply that white full-day kindergarteners outscore their half-day counterparts in both mathematics and reading. In mathematics, I estimate that full-day kindergartners have roughly a seventeen percent gain relative to their half-day peers, while in reading the corresponding advantage is nearly nineteen percent. The estimates upon which these gains are based are both very precisely estimated. Examining these short-run estimates by gender (see the top panel of

---

for short-run estimates. Each of these tables contains results from a Hausman test for the equality of full-day kindergarten coefficients in Models 2 and 3 relative to Model 1, the simple regression. See pp. 338-342 of Johnston and DiNardo (1997) for details. Finally, note that I fail to reject the equality of full-day kindergarten coefficients for Model 1 vs. Model 3 in eleven of twelve cases.

[16] Balanced and unbalanced samples yield very similar estimates and exhibit the same temporal pattern.

[17] Relevant tables also include estimated differences as a percentage of the model-specific mean.

Table 5) reveals somewhat larger gains for full-day boys relative to full-day girls in both reading and math. In reading, the estimates suggest that boys who are full-day kindergartners exhibit a nearly twenty-two percent advantage over their male half-day kindergarten counterparts, while corresponding girls have a smaller, yet still substantial, advantage of roughly fifteen percent. These boys also have larger gains in math, but the gender difference is much smaller (18.0 and 15.3 percent for boys and girls, respectively).

Longer-run estimates, contained in Table 2B, tell a much different story. For example, the seventeen percent advantage in math, reported in Table 2A, falls by more than half, to less than eight percent. The corresponding nineteen percent advantage in reading slides even further to just over five percent. While much closer to zero than short-run estimates in Table 2A, the longer-run estimates contained in Table 2B are likewise precisely estimated, though the reading coefficient in my most preferred specification should be considered only marginally significant (p-value=0.10). Hence, the gains of white full-day kindergartners, relative to their half-day peers, decline substantially over the course of an additional year so that they may now lack practical significance. To further emphasize the difference between short and longer-run estimates, consider the implied percentage gains if we assume that longer-run point estimates from my most preferred specification are, in fact, one standard error larger than reported (e.g., the mathematics full-day kindergarten coefficient in Table 2B would be 0.656+0.264=0.920). Doing this raises estimated longer-run gains to 10.8 percent and 9.0 percent in math and reading, respectively. By contrast, recall that corresponding short-run gains, without this upward adjustment, were 17.1 percent and 18.7 percent.

Examining longer-run gains by gender (see the top panel of Table 5) shows virtually no difference in the longer-run estimates in either the mathematics score (7.8 percent for girls and 8.5 percent for boys) or the reading score (6.2 percent for girls and 5.6 percent for boys). Overall, the estimated pattern for white children is one of relatively large differences by the end of kindergarten that depreciate substantially one year later, by the end of first grade. [18]

B. Black Children

Similar to, but to a lesser extent than white children, black full-day kindergartners outperform their half-day peers in both mathematics and reading by the end of kindergarten. As seen in Table 3A, black full-day kindergartners have gains in math and reading scores that are roughly eleven percent higher than their half-day counterparts. Both of these gains are estimated relatively precisely, though the reading estimate should be considered only marginally significant relative to conventional levels (p-value=0.06). Unlike white children, estimates suggest possible gender differences among black full-day kindergartners. For example, I estimate that full-day kindergarten boys exhibit a short-run mathematics score gain that is roughly sixteen percent higher than their half-day peers. Note that despite the relatively small sample size, this gain is estimated very precisely. Conversely, I find that black girls enrolled in full-day kindergarten have only a nine percent gain relative to their half-day peers. Estimated short-run reading score differences are virtually identical for black girls and boys.

---

[18] Though I said that I would omit discussion of other estimates, note that when moving from column 2 to column 3 in Tables 2 through 4, the estimated effects of family and individual covariates such as parental education and disability status are reduced in magnitude substantially. This suggests that the baseline test score is highly correlated with these characteristics. This pattern holds generally for all races. Note also, especially with respect to short-run estimates, that standard errors drop considerably. In other words, the addition of the baseline score reduces residual variance, as one might expect.

Relative to short-run estimates, longer-run results paint a much different picture. Moving from Table 3A to Table 3B, it is apparent that estimated short-run gains have largely disappeared. Once again, large percentage differences remain even if we consider them in the context of the longer-run point estimates from my most preferred specification being one standard error larger. If so, short versus longer-run percentage gains become, respectively, 11.1 percent versus 2.7 percent for math and 11.1 percent versus -0.3 percent for reading.

Further, as can be seen in the middle panel of Table 5, the sixteen percent short-run advantage enjoyed by black male full-day kindergarteners is estimated, albeit imprecisely, at less than one percent in the longer-run. So, if there are short-run gains to black boys who attend full-day kindergarten, they appear to have vanished one year later, by the end of first grade. The corresponding results for girls paint an even bleaker picture, as point estimates imply that these full-day kindergartners actually score lower than their half-day counterparts, especially in reading where estimates imply a nearly eleven percent *dis*advantage for black girls enrolled in full-day kindergarten. However, a null relationship can not be rejected in these longer-run models.

C. Hispanic Children

To a greater extent than either white or black children, I find that Hispanic full-day kindergartners outperform their half-day counterparts in both mathematics and reading in the short-run. For example, Table 4A shows that Hispanic full-day kindergartners exhibit a short-run mathematics score gain that is nearly sixteen percent higher and a corresponding reading score gain that is nearly twenty-four percent higher than their half-day counterparts. Both estimated differences are statistically different from zero at conventional levels. Given the concern over the academic achievement of Hispanic children, who tend to lag behind both white and black children, it is likely that these findings would be interpreted as "good news" by educators and policymakers alike (NCES, 2002a).[19] Examining these short-run estimates by gender (see the bottom panel of Table 5) demonstrates that full-day kindergarten tends to benefit Hispanic boys relatively more than girls in mathematics, and vice versa in reading, though differences are small.

Moving to the longer-run results reported in Table 4B, the story changes dramatically. I find no evidence that Hispanic full-day kindergartners retain any of their sizeable short-run advantages. Indeed, I find that these children actually score lower than their half-day peers one year later. While not statistically different from zero, preferred specifications imply that Hispanic full-day kindergartners have longer-run mathematics and reading score disadvantages of roughly nine percent and one percent, respectively. Note, however, that these estimated disadvantages are small, in absolute value, relative to corresponding short-run estimates. If we again compare short-run to longer-run point estimates where the latter are augmented by their own standard error, substantial differences remain. In particular, the implied mathematics difference is 15.6 percent versus -2.1 percent and the corresponding reading difference is 23.9 percent versus 7.1 percent.

Examining these longer-run gains by gender (see the bottom panel of Table 5) suggests that both Hispanic girls and boys score lower in mathematics relative to their half-day peers, though these gains are not very precisely estimated. In sum, the pattern of

---

[19] According to a recent Digest of Education Statistics, as of October 2000, the fraction of Hispanics aged sixteen to twenty-four classified as "status" dropouts was 27.8 percent, as compared to 6.9 percent of whites and 13.1 percent of blacks (NCES, 2002a).

large short-run gains, followed by much smaller longer-run differences appears most pronounced for Hispanic children.[20]

D. Investigating the Counterfactual

As noted earlier, full-day kindergartners spend about three more hours in school each day than their half-day peers. Therefore, the additional schooling inherent in full-day kindergarten is a natural explanation for the substantial short-run differences seen in Tables 2 through 4. However, it is also possible that what half-day kindergartners do with the time that they are *not* in school (and would be, if they were full-day students) is responsible for these differences. For example, if half-day kindergartners spend time in activities that are detrimental to learning, estimated differences may be driven by these activities and not the different amounts of time spent in school. Hence, it is important to understand how half-day kindergartners spend these two or three extra hours, and, more importantly, if they drive the estimated short-run differences in Tables 2 through 4.

A likely "time-use" difference between full and half-day kindergartners pertains to child care. In my analysis sample, half-day children are about ten percent more likely to be in some form of non-parental child care (51.2 percent versus 46.7 percent), but, conditional on being in such care, spend much more time per week in it (22.7 versus 8.3 hours per week) than full-day kindergartners. For a five-day school week, this translates into about 2.9 hours per day, which is roughly equivalent to the time difference between a typical full and half-day session.[21]

---

[20] Recall that I perform several sensitivity analyses of results presented in the previous three sub-sections in Tables A2A and A2B. See corresponding table notes for important details.

[21] Note, however, that ECLS-K data does not specify that child care is received only on school days. Hence, parental responses to relevant questions may include night and/or weekend hours.

Table 6 presents estimates from models that compare full-day kindergartners and their half-day peers, by whether the latter receive any non-parental child care. To be explicit, I compare (1) full-day kindergartners vs. half-day kindergartners who receive child care only from a parent, and (2) full-day kindergartners to those half-day kindergartners who receive some amount of non-parental child care.[22] So, if child care differences are driving my estimates, then these two comparisons should produce substantially different estimates of the effect of full-day kindergarten. For example, if non-parental child care is somehow more detrimental to learning, on average, than parental care, then we would expect the first comparison to result in smaller estimated differences than those presented in Tables 2 through 4 and the second comparison to generate relatively larger ones.

The upper panel of Table 6 displays results from these two comparisons as they pertain to mathematics.[23] In general, estimated coefficients for white and Hispanic children exhibit fairly small differences, though coefficients in the second row tend to be slightly further from zero than those in the first row. Coefficient estimates for black children suggest much smaller and statistically insignificant differences between full and half-day kindergartners when half-day children receive child care only from their parents. As seen in the lower panel of Table 6, however, in reading, black full-day kindergartners outscore their half-day counterparts who receive only parental care. Once again, subject-specific estimates for white and Hispanic children are similar to each other and are similar to results presented in Tables 2 and 4. Now, however, estimates in the first row

---

[22] In the context of equation (1), I produce separate estimates of γ based on whether or not half-day kindergartners receive non-parental child care.

[23] The first row of each panel (upper and lower) presents results from the comparison of full-day kindergartners and half-day kindergartners who receive no non-parental care, while the second row does so for the comparison of full-day children and half-day children who receive non-parental child care.

are slightly larger than those in the second row. So, with the possible exception of black children, it does not appear that differences in time use are driving the estimated differences between full and half-day kindergartners in Tables 2 through 4.[24]

V. Extensions

A. Distributional Impacts

While separate estimation by race gives some clues about the distributional impacts of full-day kindergarten, it provides little understanding of who benefits within a particular race group. For example, are estimated gains accruing to children with plausibly disadvantaged backgrounds or is full-day kindergarten serving to increase the gap between the better and worse off? More importantly, are there gains to disadvantaged groups that actually persist over time?

One way to understand the distributional impacts of full-day kindergarten is to estimate variants of equation (1) separately for some reasonable delineation of "disadvantaged vs. advantaged" children. Conceptually, there are many ways to characterize this distinction. For example, one might reasonably conjecture that college-educated parents are more likely to understand the value of education, broadly speaking, and hence invest in more (or better quality) time-intensive learning activities with their children (e.g., frequently read to them). This suggests estimating models separately by parents' educational attainment. Similarly, one might suspect that income plays a crucial role, since investment in children, or "child quality", is typically considered a normal good (c.f., Becker and Lewis, 1973).

[24] This conclusion relies on the assumption that the extra time half-day kindergartners spend in child care occurs when they would be in school, if they were full-day students.

While these are reasonable margins, I split my sample into "advantaged versus disadvantaged" by whether or not the child ever participated in Head Start, a national program targeted towards "at risk" children and their families, prior to entering school. In my overall analysis sample, those who had ever participated in Head Start lived in families with much lower household incomes (twenty-two thousand dollars vs. fifty-seven thousand dollars), were much less likely to have at least one parent with a college education (five percent vs. thirty-four percent) and were less likely to live in a two-parent household (fifty-one percent vs. eighty-one percent). Since black children, and to a lesser extent Hispanics, are over-represented in the Head Start population, I report these sample characteristics by race in Table 7. Regardless of race, prior Head Start participants appear much more disadvantaged than their non-Head Start peers. Hence, it provides a somewhat natural contrast to understand possible distributional impacts of full-day kindergarten.[25]

Corresponding regression estimates imply that only white Head Start participants experience greater advantages than their non-Head Start peers. More interestingly, I find evidence that these gains actually persist over time, at least to the end of first grade. With respect to mathematics, the top panel of Table 8 shows a gain that is nearly twice as large in the short-run (29.4 percent vs. 15.5 percent) and even larger, relative to non-Head Start children, in the longer-run (18.5 percent vs. 6.6 percent). A similar, though less dramatic pattern, is seen with respect to reading. Here, the longer-run gain in reading suggests some degree of persistence, though it is based on an estimated coefficient that is not even

[25] While Head Start status proxies for many dimensions of disadvantage, there is the possibility that among the disadvantaged it reflects more "motivated" families, since participation is voluntary.

marginally significant. So, among disadvantaged white children, full-day kindergarten may actually have lasting effects.

Unlike white children, I find little evidence of differences between advantaged and disadvantaged minority students. If anything, results from the middle and bottom panels of Table 8 suggest that non-Head Start children have larger short-run gains. In particular, it appears that large short-run gains accrue primarily, albeit temporarily, to relatively advantaged Hispanic children. I find no evidence of persistent gains among black or Hispanic children.

B. Why do full-day effects fade over time?

To this point, I have presented evidence that gains associated with full-day kindergarten fall off sharply just one year later, towards the end of first grade. The next natural question is why—why do these short-run effects fade? While this is a difficult question to answer, an understanding of *when* gains are lost may provide at least some information about *why* they are lost. For example, if gains are lost over the summer vacation that is common to nearly all U.S. schools, the root of the problem may be more likely linked to the child's home and/or neighborhood environment, and not directly school-related. Indeed, education researchers have documented that the gains associated with many interventions peak for a short time and then decline when the intervention is ended or otherwise disrupted (Cooper et al., 1996; Entwisle and Alexander, 1992).[26] Conversely, if gains are sustained over the summer only to be lost during the next school year, this is more consistent with a school-related explanation (e.g., lack of coordination between kindergarten and first grade curricula). This latter explanation, of course, does

[26] This phenomenon is often referred to as "summer fallback". Related findings have led to calls for year-round schooling, in the hope that achievement gains might be sustained over time.

not preclude a role for the home or neighborhood environment; indeed, the two explanations are not mutually exclusive.

I use individuals from my analysis sample who were included in the ECLS-K's randomly chosen thirty percent subsample, to understand if the short-run gains I measure have disappeared by the start of first grade, consistent with summer fallback, or if they persist over the summer, only to be lost sometime during first grade. Table 9 presents results from regressions which correspond exactly to my most preferred specification, the only exception being that the dependent variable is now measured near the start of first grade, when the thirty percent subsample was collected.

I find evidence of summer fallback for black children, but much less for Hispanic, and especially, white children. More specifically, while black full-day kindergartners show advantages of roughly eleven percent in math and reading by the end of kindergarten, comparable estimates decline to 5.6 percent and -18.5 percent, respectively, by the start of first grade. Results for Hispanic children are more mixed. While Hispanic full-day kindergarteners gain slightly in their mathematics advantage (15.6 percent to 19.3 percent), their advantage in reading slips from nearly twenty-four percent to just under eleven percent, relative to their half-day counterparts. Estimates for white children are very consistent with short-run results and hence offer even less evidence of summer fallback. In particular, I estimate that at the beginning of first grade white full-day kindergarteners have a 13.9 percent advantage over their half-day peers in math and 18.7 percent in reading, compared to 17.1 percent and 18.7 percent, respectively, at the end of kindergarten—before summer vacation.

Taken as a whole, the evidence presented suggests that the gains of black children have disappeared by the start of first grade, while the gains of white and Hispanic children apparently diminish later, during the school year. Since the short-run gains of black children are smaller than whites or Hispanics, perhaps this is not too surprising. Nevertheless, the finding is consistent with the larger literature on summer fallback (Cooper et. al., 1996). From a policy perspective, additional effort to understand the nature of these losses is likely worthwhile.

VI. Conclusions

The estimated pattern of results suggests that full-day kindergarten substantially raises the math and reading achievement of children of all races. However, these gains are much smaller in magnitude when measured via similar tests just one year later. In other words, the short-run impact of full-day kindergarten has depreciated considerably by the end of first grade. The observed pattern is even more striking for minority children since some of the specifications imply that these full-day kindergartners actually perform worse than their half-day counterparts by the end of first grade. Hispanic children, in particular, exemplify this pattern.

This pattern runs contrary to the notion that full-day kindergarten augments child human capital in a manner that allows for improved learning as children progress through school. While the estimates show substantial depreciation for all groups of children, declines are shallowest for whites. Given existing socioeconomic differences between the races, it is possible that differences in home environment contribute significantly to the larger losses for Hispanic, and especially black, children. This explanation is consistent with the summer fallback evidence I presented. Another possible explanation is that black and Hispanic children, especially public school students, are relegated to poorer quality schools that teach at a lower level, on average, than those attended by their white peers. Of course, both explanations may contribute simultaneously to the observed patterns.

While the above mentioned results paint a bleak picture of the longer-run prospects of full-day kindergarten, the estimated experience of disadvantaged white children perhaps offers some hope. Recall that among white children, estimated gains from full-day kindergarten were largest and most persistent for prior Head Start participants. Hence, at least for this particular group of plausibly disadvantaged children, full-day kindergarten may have lasting beneficial impacts. The lack of similar effects for black and Hispanic children is disappointing, especially since more traditional delineations of disadvantage (e.g., allowing the estimated effect of full-day kindergarten to differ by parental income or education) yield similar findings.

Given the limited availability of funds that can be devoted to early childhood education and the considerable costs associated with the provision of a full-day program, it is reasonable to want an improved understanding of its academic returns. The recent acceleration of the movement back to full-day kindergarten makes investigation of this topic especially germane. Finally, it should be noted that a complete evaluation of full-day kindergarten is beyond the scope of this paper. Beyond possible academic returns, issues related to child socialization and full-day kindergarten's implicit child care subsidy are also key ingredients in making a more complete assessment and hence merit future attention from researchers.

Table 1—Selected sample characteristics, by kindergarten type

	White		Black		Hispanic	
	HDK	FDK	HDK	FDK	HDK	FDK
Baseline Mathematics Score	21.03	20.79	16.45	16.39	17.10	17.43
	(7.19)	(7.14)	(5.50)	(5.12)	(6.18)	(6.03)
Baseline Reading Score	23.02	23.03	19.14	19.72	19.13	19.90
	(8.26)	(8.12)	(6.78)	(6.45)	(7.20)	(7.76)
Family Income (in 1,000s)	64.1	56.3	30.7	25.4	40.8	37.1
	(55.3)	(56.9)	(25.5)	(27.5)	(30.0)	(33.4)
Child age (in months)	68.5	68.8	67.5	68.1	67.2	68.6
	(4.1)	(4.0)	(3.9)	(4.0)	(3.9)	(4.0)
Mother's age (in years)	34.0	33.2	31.5	31.4	31.6	31.8
	(5.7)	(5.9)	(7.8)	(7.7)	(6.3)	(6.4)
At least one parent HS grad or higher	0.97	0.96	0.89	0.84	0.82	0.83
At least one parent college grad or higher	0.39	0.34	0.15	0.10	0.15	0.14
Mother in labor force	0.70	0.74	0.79	0.84	0.71	0.70
Mother works 35+ hours per week	0.41	0.47	0.57	0.58	0.46	0.49
Mom worked between birth & kindergarten	0.76	0.78	0.82	0.79	0.71	0.68
Family ever received AFDC/TANF	0.12	0.13	0.50	0.50	0.28	0.26
Parent expects kid college grad or higher	0.74	0.73	0.73	0.72	0.77	0.80
Child resides in two parent family	0.86	0.86	0.44	0.39	0.78	0.71
Ever participated in Head Start	0.07	0.10	0.38	0.45	0.21	0.22
Child disabled	0.15	0.15	0.13	0.12	0.10	0.13
Child born 2+ weeks premature	0.17	0.16	0.18	0.19	0.17	0.17
Sample Size	2,944	2,615	323	1,122	608	552

Notes: Samples consist of public school children who are first-time kindergartners and did not change schools during kindergarten. Additional restrictions make these samples correspond exactly to those that generate regression model estimates in Tables 2 through 4. The only exception is household income, which is based on slightly smaller samples (5,075 for whites, 1,216 for blacks and 1,012 for Hispanics) due to missing information. Note that differences in baseline math and reading scores between full and half-day kindergartners are not statistically different from zero at conventional levels for any race group (for whites: t=-1.26 and t=0.05, for blacks: t=-0.17 and t=1.42, and for Hispanics: t=0.92 and t=1.75, for math and reading, respectively). FDK is full-day kindergarten and HDK is half-day kindergarten. Standard deviations are in parentheses for non-binary variables.

Table 2A—Short-run regression estimates for white children

	Mathematics			Reading		
Selected covariates	1	2	3	1	2	3
Full-day kindergarten (FDK)	0.979	1.830	1.437	1.754	2.444	1.843
	(0.347)	(0.316)	(0.241)	(0.441)	(0.393)	(0.263)
	[11.7%]	[21.8%]	[17.1%]	[17.8%]	[24.8%]	[18.7%]
	{3.3%}	{6.2%}	{4.9%}	{5.3%}	{7.4%}	{5.5%}
Baseline test score	----	----	0.893	----	----	0.929
			(0.014)			(0.012)
R^2	0.005	0.22	0.66	0.01	0.20	0.67
Dependent mean/σ	29.50/8.40	29.50/8.40	29.50/8.40	33.21/9.87	33.21/9.87	33.21/9.87
Hausman test (p-values)	----	0.02	0.38	----	0.24	0.63
Sample size	5,559	5,559	5,559	5,559	5,559	5,559

Notes: Dependent variable is end-of-kindergarten test score. Sample restricted to white public school students who are first-time kindergartners and do not switch schools during kindergarten. Model 1 is a simple regression. Models 2 and 3 are identical except Model 3 includes the appropriate baseline test score. With the exception of Model 1, all models also include child gender, whether child disabled, maximum of parent education (5 categories), household income (7 categories), child age (in months), family type (4 categories), household size, ever in Head Start, mother's age (in years), mother's current work status (4 categories), whether child in day care, child birth weight (in ounces), whether child born premature, parental educational expectations (6 categories), urbanicity (7 categories), region (4 categories), number of students in kindergarten classroom (4 categories), number of days per week school meets (3 categories), number of days in school year (3 categories), whether magnet school, whether school of choice, whether parents chose residence for school-related reasons, school enrollment (6 categories), and indicators corresponding to assessment months. Each Hausman test compares the FDK coefficient of Model 2 or 3 to the Model 1 estimate; the corresponding test statistic is distributed $\chi^2(1)$. Percent of dependent standard deviation in square brackets and percent of dependent mean in curly brackets, for FDK coefficient only. Robust standard errors, adjusted for clustering at the school level, are in parentheses.

Table 2B—Longer-run regression estimates for white children

	Mathematics			Reading		
Selected covariates	1	2	3	1	2	3
Full-day kindergarten (FDK)	0.550	0.995	0.656	0.551	1.325	0.711
	(0.331)	(0.307)	(0.264)	(0.553)	(0.502)	(0.437)
	[6.5%]	[11.7%]	[7.7%]	[4.3%]	[10.3%]	[5.5%]
	{1.2%}	{2.2%}	{1.4%}	{1.0%}	{2.3%}	{1.2%}
Baseline test score	----	----	0.739	----	----	0.928
			(0.017)			(0.019)
R^2	0.003	0.19	0.48	0.004	0.18	0.46
Dependent mean/σ	45.42/8.49	45.42/8.49	45.42/8.49	57.82/12.82	57.82/12.82	57.82/12.82
Hausman test (p-values)	----	0.05	0.51	----	0.06	0.65
Sample size	5,559	5,559	5,559	5,559	5,559	5,559

Notes: Dependent variable is end-of-1st grade test score. Sample restricted to white public school students who are first-time kindergartners and do not switch schools during kindergarten. Model 1 is a simple regression. Models 2 and 3 are identical except Model 3 includes the appropriate baseline test score. With the exception of Model 1, all models also include the extensive set of covariates listed in Table 2A. Each Hausman test compares the FDK coefficient of Model 2 or 3 to the Model 1 estimate; the corresponding test statistic is distributed $\chi^2(1)$. Percent of dependent standard deviation in square brackets and percent of dependent mean in curly brackets, for FDK coefficient only. Robust standard errors, adjusted for clustering at the school level, are in parentheses.

Table 3A—Short-run regression estimates for black children

	Mathematics			Reading		
Selected covariates	1	2	3	1	2	3
Full-day kindergarten (FDK)	0.947	1.214	0.842	1.423	1.636	0.996
	(0.530)	(0.466)	(0.350)	(0.688)	(0.626)	(0.534)
	[13.0%]	[16.7%]	[11.1%]	[15.8%]	[18.2%]	[11.1%]
	{4.0%}	{5.2%}	{3.6%}	{5.0%}	{5.7%}	{3.5%}
Baseline test score	----	----	1.011	----	----	0.990
			(0.031)			(0.036)
R^2	0.008	0.22	0.61	0.01	0.22	0.62
Dependent mean/σ	23.40/7.26	23.40/7.26	23.40/7.26	28.52/8.99	28.52/8.99	28.52/8.99
Hausman test (p-values)	----	0.54	0.81	----	0.71	0.45
Sample size	1,445	1,445	1,445	1,445	1,445	1,445

Notes: Dependent variable is end-of-kindergarten test score. Sample restricted to black public school students who are first-time kindergartners and do not switch schools during kindergarten. Model 1 is a simple regression. Models 2 and 3 are identical except Model 3 includes the appropriate baseline test score. With the exception of Model 1, all models also include the extensive set of covariates listed in Table 2A. Each Hausman test compares the FDK coefficient of Model 2 or 3 to the Model 1 estimate; the corresponding test statistic is distributed $\chi^2(1)$. Percent of dependent standard deviation in square brackets and percent of dependent mean in curly brackets, for FDK coefficient only. Robust standard errors, adjusted for clustering at the school level, are in parentheses.

Table 3B—Longer-run regression estimates for black children

	Mathematics			Reading		
Selected covariates	1	2	3	1	2	3
Full-day kindergarten (FDK)	-0.177	0.092	-0.312	-0.099	-0.266	-1.044
	(0.701)	(0.625)	(0.544)	(1.050)	(1.111)	(1.003)
	[-2.0%]	[1.1%]	[-3.6%]	[-0.7%]	[-2.0%]	[-7.8%]
	{-0.5%}	{0.2%}	{-0.8%}	{-0.2%}	{-0.5%}	{-2.1%}
Baseline test score	----	----	1.041	----	----	1.171
			(0.044)			(0.063)
R^2	0.003	0.21	0.51	0.007	0.19	0.45
Dependent mean/σ	38.70/8.69	38.70/8.69	38.70/8.69	50.36/13.32	50.36/13.32	50.36/13.32
Hausman test (p-values)	----	0.62	0.80	----	0.84	0.25
Sample size	1,445	1,445	1,445	1,445	1,445	1,445

Notes: Dependent variable is end-of-1st grade test score. Sample restricted to black public school students who are first-time kindergartners and do not switch schools during kindergarten. Model 1 is a simple regression. Models 2 and 3 are identical except Model 3 includes the appropriate baseline test score. With the exception of Model 1, all models also include the extensive set of covariates listed in Table 2A. Each Hausman test compares the FDK coefficient of Model 2 or 3 to the Model 1 estimate; the corresponding test statistic is distributed $\chi^2(1)$. Percent of dependent standard deviation in square brackets and percent of dependent mean in curly brackets, for FDK coefficient only. Robust standard errors, adjusted for clustering at the school level, are in parentheses.

Table 4A—Short-run regression estimates for Hispanic children

	Mathematics			Reading		
Selected covariates	1	2	3	1	2	3
Full-day kindergarten (FDK)	1.381	1.237	1.219	2.328	3.155	2.278
	(0.578)	(0.605)	(0.436)	(0.695)	(0.810)	(0.623)
	[17.7%]	[15.8%]	[15.6%]	[24.4%]	[33.1%]	[23.9%]
	{5.4%}	{4.9%}	{4.8%}	{7.7%}	{10.5%}	{7.5%}
Baseline test score	----	----	0.957	----	----	0.937
			(0.025)			(0.029)
R^2	0.02	0.27	0.67	0.03	0.23	0.63
Dependent mean/σ	25.49/7.81	25.49/7.81	25.49/7.81	30.18/9.52	30.18/9.52	30.18/9.52
Hausman test (p-values)	----	0.77	0.74	----	0.15	0.93
Sample size	1,160	1,160	1,160	1,160	1,160	1,160

Notes: Dependent variable is end-of-kindergarten test score. Sample restricted to Hispanic public school students who are first-time kindergartners and do not switch schools during kindergarten. Model 1 is a simple regression. Models 2 and 3 are identical except Model 3 includes the appropriate baseline test score. With the exception of Model 1, all models also include the extensive set of covariates listed in Table 2A. Each Hausman test compares the FDK coefficient of Model 2 or 3 to the Model 1 estimate; the corresponding test statistic is distributed $\chi^2(1)$. Percent of dependent standard deviation in square brackets and percent of dependent mean in curly brackets, for FDK coefficient only. Robust standard errors, adjusted for clustering at the school level, are in parentheses.

Table 4B—Longer-run regression estimates for Hispanic children

	Mathematics			Reading		
Selected covariates	1	2	3	1	2	3
Full-day kindergarten (FDK)	0.695	-0.633	-0.730	0.961	0.826	-0.141
	(0.578)	(0.646)	(0.554)	(0.695)	(1.139)	(1.064)
	[8.3%]	[-7.6%]	[-8.7%]	[7.4%]	[6.4%]	[-1.1%]
	{1.7%}	{-1.5%}	{-1.8%}	{1.8%}	{1.5%}	{-0.3%}
Baseline test score	----	----	0.802	----	----	0.997
			(0.039)			(0.050)
R^2	0.007	0.22	0.47	0.005	0.20	0.45
Dependent mean/σ	41.94/8.37	41.94/8.37	41.94/8.37	53.46/12.98	53.46/12.98	53.46/12.98
Hausman test (p-values)	----	0.01	0.02	----	0.87	0.19
Sample size	1,160	1,160	1,160	1,160	1,160	1,160

Notes: Dependent variable is end-of-1st grade test score. Sample restricted to Hispanic public school students who are first-time kindergartners and do not switch schools during kindergarten. Model 1 is a simple regression. Models 2 and 3 are identical except Model 3 includes the appropriate baseline test score. With the exception of Model 1, all models also include the extensive set of covariates listed in Table 2A. Each Hausman test compares the FDK coefficient of Model 2 or 3 to the Model 1 estimate; the corresponding test statistic is distributed $\chi^2(1)$. Percent of dependent standard deviation in square brackets and percent of dependent mean in curly brackets, for FDK coefficient only. Robust standard errors, adjusted for clustering at the school level, are in parentheses.

Table 5—Full-day kindergarten coefficient estimates, by race and gender.

	Mathematics				Reading			
	Short-run		Longer-run		Short-run		Longer-run	
	Males	Females	Males	Females	Males	Females	Males	Females
White children								
Full-day kindergarten (FDK)	1.587	1.212	0.759	0.626	2.151	1.487	0.740	0.761
	(0.341)	(0.256)	(0.350)	(0.314)	(0.329)	(0.341)	(0.544)	(0.515)
	[18.0%]	[15.3%]	[8.5%]	[7.8%]	[21.5%]	[15.4%]	[5.6%]	[6.2%]
	{5.4%}	{4.1%}	{1.7%}	{1.4%}	{6.6%}	{4.4%}	{1.3%}	{1.3%}
R^2	0.67	0.65	0.49	0.50	0.68	0.66	0.46	0.48
Sample size	2,829	2,730	2,829	2,730	2,829	2,730	2,829	2,730
Black children								
Full-day kindergarten (FDK)	1.188	0.652	0.061	-0.463	0.889	0.903	-0.223	-1.367
	(0.479)	(0.467)	(0.703)	(0.749)	(0.806)	(0.621)	(1.478)	(1.311)
	[15.8%]	[9.3%]	[0.7%]	[-5.6%]	[9.6%]	[10.4%]	[-1.6%]	[-10.7%]
	{5.1%}	{2.8%}	{0.2%}	{-1.2%}	{3.2%}	{3.1%}	{-0.5%}	{-2.6%}
R^2	0.68	0.57	0.53	0.52	0.65	0.62	0.49	0.44
Sample size	709	736	709	736	709	736	709	736
Hispanic children								
Full-day kindergarten (FDK)	1.404	0.901	-1.276	-0.416	1.869	2.621	-0.236	-0.555
	(0.520)	(0.631)	(0.872)	(0.700)	(0.709)	(0.790)	(1.344)	(1.254)
	[18.4%]	[11.3%]	[-14.8%]	[-5.1%]	[20.3%]	[27.0%]	[-1.8%]	[-4.4%]
	{5.6%}	{3.5%}	{-3.0%}	{-1.0%}	{6.5%}	{8.3%}	{-0.5%}	{-1.0%}
R^2	0.69	0.69	0.49	0.53	0.64	0.67	0.45	0.53
Sample size	594	566	594	566	594	566	594	566

Notes: Samples restricted to public school students who are first-time kindergartners and do not switch schools during kindergarten. All models include the appropriate baseline test score and the extensive set of covariates listed in Table 2A, with the exception of child gender. Percent of dependent standard deviation in square brackets and percent of dependent mean in curly brackets. Robust standard errors, adjusted for clustering at the school level, are in parentheses.

Table 6—Full-day kindergartners versus half-day kindergartners, by whether or not the latter receive non-parental child care

	White		Black		Hispanic	
	Short-run	Longer-run	Short-run	Longer-run	Short-run	Longer-run
Dependent variable: Math score						
Full-day kindergarten (FDK)	1.275	0.574	0.571	-0.717	1.387	-0.573
(vs. half-day kindergarteners who	(0.235)	(0.298)	(0.508)	(0.829)	(0.578)	(0.710)
receive no non-parental child care)	[4,072]	[4,072]	[1,247]	[1,247]	[855]	[855]
Full-day kindergarten (FDK)	1.445	0.785	1.128	-0.108	1.060	-0.941
(vs. half-day kindergarteners who	(0.281)	(0.305)	(0.382)	(0.603)	(0.447)	(0.683)
receive non-parental child care)	[4,089]	[4,089]	[1,317]	[1,317]	[855]	[855]
Dependent variable: Reading score						
Full-day kindergarten (FDK)	1.907	0.645	1.318	-1.529	2.178	0.388
(vs. half-day kindergarteners who	(0.303)	(0.507)	(0.669)	(1.381)	(0.768)	(1.248)
receive no non-parental child care)	[4,072]	[4,072]	[1,247]	[1,247]	[855]	[855]
Full-day kindergarten (FDK)	1.517	0.602	0.931	-0.904	1.948	-0.762
(vs. half-day kindergarteners who	(0.300)	(0.470)	(0.598)	(1.130)	(0.686)	(1.163)
receive non-parental child care)	[4,089]	[4,089]	[1,317]	[1,317]	[855]	[855]

Notes: Samples restricted to public school students who are first-time kindergartners and do not switch schools during kindergarten. All models include the appropriate baseline test score and the extensive set of covariates listed in Table 2A, with the exception of the child care variable. Sample sizes in brackets. Robust standard errors, adjusted for clustering at the school level, are in parentheses.

Table 7—Selected sample characteristics, by prior Head Start participation status

	White		Black		Hispanic	
Selected Characteristics	HS	No HS	HS	No HS	HS	No HS
Household Income (in 1000s)	27.0	63.5	18.5	32.6	22.4	43.6
At least one parent has college degree or higher	0.07	0.40	0.05	0.16	0.03	0.18
Child resides in two parent family	0.72	0.88	0.33	0.45	0.60	0.78
Sample size	470	5,089	815	630	250	910

Notes: Samples correspond to analysis samples as outlined in section III (N=5,559 for whites, N=1,445 for blacks and N=1,160 for Hispanics). As in Table 1, household income is based on fewer observations than other covariates listed. HS represents those who have ever participated in Head Start.

Table 8—Short and longer-run estimates of the effect of full-day kindergarten, by prior Head Start participation status.

	Mathematics				Reading			
	Short-run		Longer-run		Short-run		Longer-run	
	HS	Non-HS	HS	Non-HS	HS	Non-HS	HS	Non-HS
White children								
Full-day kindergarten (FDK)	2.447	1.289	1.776	0.547	1.364	1.766	1.667	0.499
	(0.701)	(0.245)	(0.905)	(0.268)	(0.808)	(0.271)	(1.516)	(0.439)
	[29.4%]	[15.5%]	[18.5%]	[6.6%]	[15.0%]	[18.0%]	[12.4%]	[4.0%]
	{9.6%}	{4.3%}	{4.4%}	{1.2%}	{4.7%}	{5.3%}	{3.3%}	{0.9%}
R^2	0.74	0.65	0.58	0.47	0.69	0.66	0.52	0.45
Sample size	470	5,089	470	5,089	470	5,089	470	5,089
Black children								
Full-day kindergarten (FDK)	0.872	0.639	-0.733	0.151	0.657	1.110	-1.154	-1.316
	(0.481)	(0.464)	(0.773)	(0.741)	(0.650)	(0.656)	(1.330)	(1.180)
	[13.0%]	[8.4%]	[-8.8%]	[1.7%]	[8.3%]	[11.5%]	[-9.3%]	[-9.5%]
	{3.9%}	{2.7%}	{-1.9%}	{0.4%}	{2.4%}	{3.8%}	{-2.4%}	{-2.5%}
R^2	0.62	0.63	0.48	0.55	0.60	0.64	0.45	0.47
Sample size	630	815	630	815	630	815	630	815
Hispanic children								
Full-day kindergarten (FDK)	-0.405	1.443	-2.447	-0.356	0.767	2.525	-2.965	0.365
	(0.894)	(0.474)	(1.182)	(0.676)	(1.306)	(0.619)	(2.369)	(1.057)
	[-5.8%]	[18.3%]	[-29.9%]	[-4.3%]	[8.9%]	[26.2%]	[-23.0%]	[2.8%]
	{-1.8%}	{5.5%}	{-6.2%}	{-0.8%}	{2.8%}	{8.2%}	{-5.9%}	{0.7%}
R^2	0.71	0.67	0.59	0.46	0.66	0.65	0.48	0.48
Sample size	250	910	250	910	250	910	250	910

Notes: Samples restricted to public school students who are first-time kindergartners and do not switch schools during kindergarten. All models include the appropriate baseline test score and the extensive set of covariates listed in Table 2A, with the exception of the prior Head Start participation variable. Percent of dependent standard deviation in square brackets and percent of dependent mean in curly brackets. Robust standard errors, adjusted for clustering at the school level, are in parentheses. HS includes those children who have ever participated in Head Start.

Table 9—Beginning of 1st grade regression estimates

	Mathematics			Reading		
Selected covariates	White	Black	Hispanic	White	Black	Hispanic
Full-day kindergarten (FDK)	1.260	0.471	1.663	2.311	-1.930	1.255
	(0.411)	(0.963)	(1.034)	(0.531)	(1.099)	(1.173)
	[13.9%]	[5.6%]	[19.3%]	[18.7%]	[-18.5%]	[10.7%]
	{3.6%}	{1.7%}	{5.4%}	{5.8%}	{-5.7%}	{3.4%}
Baseline test score	0.916	1.127	0.905	1.095	1.062	1.002
	(0.026)	(0.068)	(0.068)	(0.027)	(0.065)	(0.066)
R^2	0.65	0.66	0.68	0.64	0.67	0.72
Dependent mean/σ	34.96/9.07	28.27/8.41	30.85/8.61	40.11/12.33	33.96/10.46	36.49/11.72
Sample size	1,620	392	312	1,620	392	312

Notes: Dependent variable is start-of-1st grade test score. Samples restricted to public school students who are first-time kindergartners and do not switch schools during kindergarten. Models also include the extensive set of covariates listed in Table 2A. Percent of dependent standard deviation in square brackets and percent of dependent mean in curly brackets, for FDK coefficient only. Robust standard errors, adjusted for clustering at the school level, are in parentheses.

Table A1—Difference-in differences estimates versus estimates from Equation (1) with $\alpha=\delta=\theta=0$ restriction.

	Difference-in-differences estimates				Equation (1) with $\alpha=\delta=\theta=0$ restriction			
	Math		Reading		Math		Reading	
	SR	LR	SR	LR	SR	LR	SR	LR
White children (N=5,559)								
Full-day kindergarten (FDK)	1.105	0.787	1.606	0.499	1.089	0.737	1.607	0.499
	(0.194)	(0.227)	(0.249)	(0.388)	(0.194)	(0.222)	(0.250)	(0.388)
	[13.2%]	[9.3%]	[16.3%]	[3.9%]	[13.0%]	[8.7%]	[16.3%]	[3.9%]
	{3.7%}	{1.7%}	{4.8%}	{0.9%}	{3.7%}	{1.6%}	{4.8%}	{0.9%}
Black children (N=1,445)								
Full-day kindergarten (FDK)	0.941	-0.177	0.776	-0.809	0.944	-0.171	0.756	-0.969
	(0.326)	(0.530)	(0.447)	(0.800)	(0.325)	(0.526)	(0.448)	(0.788)
	[13.0%]	[-2.0%]	[8.6%]	[-6.1%]	[13.0%]	[-2.0%]	[8.4%]	[-7.3%]
	{4.0%}	{-0.5%}	{2.7%}	{-1.6%}	{4.0%}	{-0.4%}	{2.7%}	{-1.9%}
Hispanic children (N=1,160)								
Full-day kindergarten (FDK)	0.944	0.300	1.431	0.143	0.944	0.343	1.461	0.091
	(0.363)	(0.467)	(0.499)	(0.721)	(0.363)	(0.471)	(0.498)	(0.723)
	[12.1%]	[3.6%]	[15.0%]	[1.1%]	[12.1%]	[4.1%]	[15.3%]	[0.7%]
	{3.7%}	{0.7%}	{4.7%}	{0.3%}	{3.7%}	{0.8%}	{4.8%}	{0.2%}

Notes: Samples restricted to public school students who are first-time kindergartners and do not switch schools during kindergarten. The difference-in-differences estimates are derived from the following regression: Δ(Test score) = $\beta_0 + \beta_1$(FDK indicator). The other set of estimates is derived from Equation (1) with the restriction that $\alpha=\delta=\theta=0$, which is equivalent to the following specification: (Outcome Test Score)$_{t+j} = \alpha + \psi$(Baseline Test Score)$_t + \gamma$(FDK indicator). Percent of outcome test score standard deviation in square brackets and percent of outcome test score mean in curly brackets. SR means short-run and LR means longer-run. Robust standard errors, adjusted for clustering at the school level, are in parentheses.

Table A2A—Checking the sensitivity of FDK estimates in Tables 2-4 for mathematics scores.

	White		Black		Hispanic	
Sample/Specification	SR	LR	SR	LR	SR	LR
Tables 2-4	1.437	0.656	0.842	-0.312	1.219	-0.730
(Repeated)	(0.241)	(0.264)	(0.350)	(0.544)	(0.436)	(0.554)
	[5,559]	[5,559]	[1,445]	[1,445]	[1,160]	[1,160]
Tables 2-4	1.151	0.681	0.814	-0.189	1.360	-0.320
(W/o school covariates)	(0.221)	(0.253)	(0.344)	(0.547)	(0.416)	(0.516)
	[5,559]	[5,559]	[1,445]	[1,445]	[1,160]	[1,160]
Tables 2-4	1.473	0.848	0.801	-0.010	1.027	-0.623
(Same school K to 1st)	(0.249)	(0.281)	(0.384)	(0.622)	(0.471)	(0.562)
	[5,084]	[5,084]	[1,287]	[1,287]	[1,039]	[1,039]
Tables 2-4	1.408	0.676	0.976	-0.174	1.114	-0.442
(All or nothing schools)	(0.256)	(0.285)	(0.368)	(0.565)	(0.471)	(0.578)
	[5,176]	[5,176]	[1,397]	[1,397]	[1,115]	[1,115]
Tables 2-4	1.331	----	0.875	----	1.060	----
(Unbalanced samples)	(0.226)		(0.296)		(0.408)	
	[6,370]		[1,739]		[1,430]	
Complete case analysis	1.388	0.617	1.095	0.010	2.142	-0.391
(With school covariates)	(0.291)	(0.311)	(0.521)	(0.675)	(0.623)	(0.749)
	[4,189]	[4,189]	[861]	[861]	[684]	[684]
Complete case analysis	1.088	0.664	0.962	-0.140	1.261	-0.533
(W/o school covariates)	(0.217)	(0.259)	(0.352)	(0.578)	(0.431)	(0.548)
	[5,075]	[5,075]	[1,216]	[1,216]	[1,012]	[1,012]
Complete case analysis	1.306	0.671	1.197	0.398	2.128	0.084
(With school & teacher	(0.303)	(0.318)	(0.572)	(0.674)	(0.682)	(0.753)
covariates)	[3,550]	[3,550]	[709]	[709]	[563]	[563]

Notes: Each row presents estimates of the effect of full-day kindergarten on the relevant mathematics score for a different sample or model specification. The first row is information that is repeated from Column 3 (my preferred specification) of Tables 2-4. The second row presents corresponding estimates from models that do not include school-level covariates. The third row presents estimates using only those children who remain in the same school between kindergarten and first grade. The fourth row provides estimates using only those schools with either all full-day or all half-day students. The fifth row presents estimates that correspond to those in Tables 2-4, but these estimates are generated with unbalanced samples. The sixth row presents estimates that correspond to Tables 2-4, but uses complete case analysis, instead of including indicators for selected missing covariates. The seventh row presents estimates that correspond to the sixth row with the exception that models do not include school covariates. Finally, the eighth row presents estimates that correspond to the seventh row, but with the following teacher characteristics added to the model: teaching experience in elementary school, teacher education level, and teacher certification type. SR stands for short-run and LR means longer-run. Sample sizes in brackets. Robust standard errors, adjusted for clustering at the school level, are in parentheses.

Table A2B—Checking the sensitivity of FDK estimates in Tables 2-4 for reading scores.

	White		Black		Hispanic	
Sample/Specification	SR	LR	SR	LR	SR	LR
Tables 2-4 (Repeated)	1.843 (0.263) [5,559]	0.711 (0.437) [5,559]	0.996 (0.534) [1,445]	-1.044 (1.003) [1,445]	2.278 (0.623) [1,160]	-0.141 (1.064) [1,160]
Tables 2-4 (W/o school covariates)	1.624 (0.262) [5,559]	0.737 (0.429) [5,559]	1.182 (0.515) [1,445]	-1.151 (0.984) [1,445]	2.392 (0.585) [1,160]	0.250 (0.965) [1,160]
Tables 2-4 (Same school K to 1st)	1.957 (0.275) [5,084]	0.636 (0.464) [5,084]	1.030 (0.564) [1,287]	-0.584 (1.085) [1,287]	2.350 (0.640) [1,039]	0.052 (1.044) [1,039]
Tables 2-4 (All or nothing schools)	1.859 (0.290) [5,176]	0.719 (0.469) [5,176]	1.027 (0.570) [1,397]	-1.056 (1.080) [1,397]	2.196 (0.670) [1,115]	-0.200 (1.162) [1,115]
Tables 2-4 (Unbalanced samples)	1.790 (0.252) [6,370]	----	0.837 (0.485) [1,739]	----	2.525 (0.561) [1,430]	----
Complete case analysis (With school covariates)	2.098 (0.321) [4,189]	0.618 (0.469) [4,189]	1.022 (0.626) [861]	-1.768 (1.348) [861]	3.116 (0.762) [684]	0.994 (1.492) [684]
Complete case analysis (W/o school covariates)	1.654 (0.265) [5,075]	0.729 (0.424) [5,075]	1.354 (0.512) [1,216]	-1.655 (1.061) [1,216]	2.305 (0.626) [1,012]	0.171 (1.021) [1,012]
Complete case analysis (With school & teacher covariates)	2.033 (0.341) [3,550]	0.569 (0.477) [3,550]	1.301 (0.684) [709]	-1.103 (1.449) [709]	2.985 (0.763) [563]	1.327 (1.540) [563]

Notes: Each row presents estimates of the effect of full-day kindergarten on the relevant reading score for a different sample or model specification. The first row is information that is repeated from Column 3 (my preferred specification) of Tables 2-4. The second row presents corresponding estimates from models that do not include school-level covariates. The third row presents estimates using only those children who remain in the same school between kindergarten and first grade. The fourth row provides estimates using only those schools with either all full-day or all half-day students. The fifth row presents estimates that correspond to those in Tables 2-4, but these estimates are generated with unbalanced samples. The sixth row presents estimates that correspond to Tables 2-4, but uses complete case analysis, instead of including indicators for selected missing covariates. The seventh row presents estimates that correspond to the sixth row with the exception that models do not include school covariates. Finally, the eighth row presents estimates that correspond to the seventh row, but with the following teacher characteristics added to the model: teaching experience in elementary school, teacher education level, and teacher certification type. SR stands for short-run and LR means longer-run. Sample sizes in brackets. Robust standard errors, adjusted for clustering at the school level, are in parentheses.

References

Becker, G.S. and H.G. Lewis (1973). On the interaction between child quantity and the quality of children. *Journal of Political Economy,* 81:2(Part 2): 279-288.

Carr, M. and Jessup, D.L. (1997). Gender differences in first grade mathematics strategy use: Social and metacognitive influences. *Journal of Educational Psychology*, 98(2): 318-328.

Cascio, E.U. (2003). Getting a head start or falling further behind? Black-white relative schooling attainment and the introduction of kindergartens in the South. Working paper.

Cooper, H. Nye, B., Charlton, K., Lindsey, J and S. Greathouse (1996). The effect of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Education Research*, 66(3): 227-268.

Cryan, J.R. et al. (1992). Successful outcomes of full-day kindergarten: More positive behavior and increased achievement in the years after. *Early Childhood Research Quarterly,* 7(2): 187-203.

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

Currie J. and D. Thomas (1999). Does Head Start help Hispanic children? *Journal of Public Economics*, 74(2): 235-262.

Deaton, A. (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, Published for the World Bank, The Johns Hopkins University Press, Baltimore and London.

Elicker, J and S. Mathur (1997). What do they do all day? Comprehensive evaluation of a full-day kindergarten. *Early Childhood Research Quarterly,* 12(4): 459-480.

Entwisle, D.R. and K. Alexander (1992). Summer setback: Race, poverty, school composition and math achievement in the first two years of school. *American Sociological Review,* 53(Feb.): 72-84.

Fennema, E. and J. Sherman (1977). Sex-related differences in mathematics achievement, spatial visualization, and affective factors. *American Educational Research Journal,* 19(1): 51-71.

Fromberg, D.P. (1992). Implementing full-day kindergarten. *Principal*, 71(5): 26-28.

Fusaro, J.A. (1997). The effect of full-day kindergarten on student achievement: A meta-analysis. *Child Study Journal* 27(4): 269-277.

Garces, E., Thomas, D. and Currie, J. (2002). Longer-term effects of Head Start. *American Economic Review*, 92(4): 999-1012.

Goldin, C. (1991). The role of World War II in the rise of women's work. *American Economic Review,* 81(4): 741-756.

Gorton, H.B. (1968). A study of the kindergarten program: Full or half day? ERIC Document 012 327.

Gullo, D.F. (1990). The changing family context: Implications for the development of all-day kindergarten. *Young Children* 45(4): 35-39.

Gullo, D.F. (2000). Long-term educational effects of half-day vs. full-day kindergarten. *Early Child Development and Care*. 160: 17-24.

Hough, D and S. Bryde (1996). The effects of full-day kindergarten on student's academic performance, Paper presented at the 1996 American Education Research Association meetings.

Housden, T. and R. Kam (1992). Full-day kindergarten: A summary of the research. ERIC Document 345 868.

Johnston, J. and J.E. DiNardo (1997). Econometric Methods. 4th Edition. New York: McGraw-Hill.

Jones, S.S. (2002). Effect of all-day kindergarten on student cognitive growth: A meta-analysis. Doctoral dissertation, University of Kansas.

Karweit, N. (1992). The kindergarten experience. *Educational Leadership* 49(6): 82-86.

Krueger, A.B. and P. Zhu (2002). Another look at the New York City school voucher experiment. NBER Working Paper, #9418.

Lemke, R.J., Witte, A.D., Querault, M. and R. Witt (2000). Child care and the welfare to work transition. NBER Working Paper #7583.

Lord, F.M. (1980). Applications of Item Response Theory to Practical Testing Problems. Hillsdale, NJ: Lawrence Erlbaum Publishers.

Moulton (1990). An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *Review of Economics and Statistics* 72(2): 334-338.

Natale, J.A. (2001). Early learners: Are full-day academic kindergartens too much, too soon? *American School Board Journal*, 188(3): 22-25.

Oelerich, M.C. (1984). Should kindergarten children attend school all day every day? *The Journal of College of Education*, Fall: 13-16.

Ohio State Legislative Office of Education Oversight (1997). *An overview of full-day kindergarten*. Columbus, OH.

Olsen, D and E. Zigler (1989). An assessment of the all-day kindergarten movement. *Early Childhood Research Quarterly*, 4(2): 167-186.

Siefert, K. (1993). Cognitive development and early childhood education. In D. Spodeck (Ed.) *Handbook of Research on the Education of Young Children.* New York: McMillan.

Ulrey et al. (1982). Effects of length of school day on kindergarten performance and parent satisfaction. *Psychology in the Schools*, 19, 238-242.

U.S. Census Bureau (2002). *The Population Profile of the United States: 2000 (Internet Release)*, www.census.gov/population/www/pop-profile/profile2000.html, Washington, D.C.

U.S. Department of Education, National Center for Education Statistics (2002a). *Digest of Education Statistics 2001*, NCES 2002-130, by Thomas Snyder. Production Manager, Charlene M. Hoffman. Washington, D.C.

U.S. Department of Education, National Center for Education Statistics (2002b). *User's Manual for the ECLS-K First Grade Public-Use Data Files and Electronic Codebook*, NCES 2002-135, Washington, D.C.

U.S. Department of Education, National Center for Education Statistics (2002c). *Early Childhood Longitudinal Study—Kindergarten Class of 1998-99 (ECLS-K), Psychometric Report For Kindergarten through First Grade*, NCES 2002-05, by Donald A. Rock and Judith M. Pollack, Educational Testing Service, Elvira Germino Hausken, project officer. Washington, D.C.

CHAPTER 4

LOCAL LABOR MARKET FLUCTUATIONS AND HEALTH: IS THERE A CONNECTION AND FOR WHOM?

I. Introduction and Background

Economists have devoted much attention to the impact of macroeconomic fluctuations on a variety of outcomes, including earnings and their distribution, criminal activity and human capital investment. Collectively, they have paid less attention to a possible connection to health. Using repeated cross sectional data from the National Health Interview Surveys (NHIS), I estimate relationships between local labor market conditions and several measures of health and health behaviors for a sample of individuals living in the fifty-eight largest metropolitan statistical areas (MSAs) in the United States.

In the remainder of this section, I discuss why health may vary with local labor market fluctuations, whose health might be most affected and the relevant literature. In Section II, I describe my data, focusing on key variables and the construction of my analysis sample, which consists of working-aged men. Section III presents my empirical strategy which relates local labor market conditions, via MSA-level unemployment rates, to measures of health and health behaviors that may vary over short periods of time. Since the effect of labor market conditions on health may depend on the extent to which one's present or prospective employment is impacted by them, I divide my sample into groups whose employment prospects are potentially more and less likely to be affected by such fluctuations. In particular, I allow the effect of local labor market conditions to vary by race and education groups since previous research suggests the labor market outcomes of non-white and less educated individuals are relatively more affected by economic fluctuations. In addition, I allow this effect to vary by one's potential "exposure" to labor market fluctuations, as measured by their predicted employment status. Section IV presents my findings. For those men least likely to be employed, I find consistent evidence of a procyclical relationship for body weight and psychological well-being, but no systematic relationship for a variety of health behaviors including cigarette smoking, heavy alcohol consumption, and various forms of physical exercise. Consistent with these findings, I present evidence that worsening labor market conditions lead to weight gains and reduced psychological well-being among African American men and lower psychological well-being among less educated males. I find scant evidence of a relationship between local labor market conditions and health behaviors, especially those related to physical exercise, for any race or education groups. Section V discusses my findings and how they compare with existing evidence. Section VI concludes the paper.

A. Why might local labor market conditions affect health?

Conceptually, local labor market conditions may affect health for a variety of, sometimes conflicting, reasons. Two general explanations have gained prominence in recent related work. To elaborate, I briefly consider each in the context of a labor market contraction.

First, local labor market fluctuations might impact health through changes in the opportunity cost of time. When the unemployment rate rises, employment is reduced on intensive and extensive margins. Such reductions lower the opportunity cost of other, non-market activities including household production. One form of household production that is very time-intensive is the production of health.[1] Facing lower time costs, affected individuals may spend more time in activities intended to improve their health (e.g., exercising, producing and consuming homemade rather than mass-produced or restaurant meals, or using preventive medical services). If investment in such activities actually improves health and does so in a reasonably short period of time, a countercyclical relationship between labor market conditions and health will obtain.[2]

Another channel through which labor market conditions might affect health is sometimes referred to as the "economic stress" hypothesis (c.f., Catalano and Dooley, 1983; Catalano, 1991). Generally speaking, the idea is that a weaker economy leads to increased stress due to greater uncertainty of present and future income receipt. Moreover, such uncertainty may increase the likelihood of stressful life events such as bankruptcy or marital dissolution which, in turn, may add to the stress associated with a downturn in the labor market. If the stress hypothesis is operative and if greater stress reduces health in the short-run, a procyclical relationship between labor market conditions and health will obtain.[3]

---

[1] As anyone who has ever purchased a piece of exercise equipment or paid dues to a health club knows, investment in health can also be quite goods-intensive, but inherently involves a substantial time component.

[2] Of course, reductions in the opportunity cost of market time make time spent in other, potentially health-reducing activities less costly as well (e.g., late nights spent at a local tavern).

[3] However, it is plausible that economic contractions reduce work-related stress. For example, it is likely that mandatory overtime and, more generally, worker effort fall during labor market contractions.

B. Who might be most affected by labor market fluctuations?

While these two general explanations are not mutually exclusive and do not exhaust the mechanisms through which labor market conditions may affect health, they do indicate that their directional impact is an empirical question. A separate issue is whose health is most likely to be impacted by such fluctuations.

Since the question of interest is whether labor market conditions impact health, individuals whose employment prospects are most affected by labor market fluctuations may be most likely to experience corresponding changes in health, if such effects exist. Consider this possibility in the context of a labor market expansion. As a group, individuals seeking employment should have improved prospects when the unemployment rate falls, while those who would be relatively more likely to lose an existing job during a contraction should face a lower probability of job loss as the economy strengthens. For such individuals, changes in the value of time or stress levels should be relatively larger than those further from the extensive employment margin. If so, and if these mechanisms are operative, we would expect to see relatively larger health effects among such individuals.

Note, however, that individuals further from the extensive employment margin also may be affected by labor market fluctuations. For example, for those who remain without jobs in an expansion, government programs that provide cash or other in-kind benefits are less likely to expire or otherwise be curtailed (e.g., unemployment insurance, job training, etc.). On the other end of the spectrum, those with relatively secure employment may be impacted since such individuals may experience improved job mobility and/or job characteristics (e.g., higher real wages or more generous fringe benefits) in an expansion. So, while labor market conditions may affect health across a wide range of individuals, it seems likely that individuals whose labor market fortunes are most impacted by fluctuations will experience the largest health effects, if they exist.

But who are these individuals? That is, who are the individuals whose current employment or employment prospects are likely to be most affected by fluctuations in local labor market conditions? Previous work on the distributional consequences of economic shocks suggests the labor market outcomes of "lower-skilled" individuals are disproportionately affected.[4] Of these studies, the ones that use MSA-level variation in labor market conditions to examine labor market outcomes such as earnings and employment are most relevant to this study (Bartik, 1991, 1993a, 1993b, 1994, 1996; Bound and Holzer, 1993, 1995). Generally speaking, these studies find greater sensitivity to economic fluctuations among non-whites, younger individuals and those with lower education levels.[5] More recently, and in one of the most comprehensive studies in this literature, Hoynes (2000) shows that the labor market outcomes of non-whites and those with lower levels of education are relatively more impacted by changes in local labor market conditions.[6] In particular, she finds that these groups are more likely to experience reductions in employment and earnings in a contraction, and more likely to experience gains in these areas in subsequent recoveries, relative to their white and more educated counterparts. Based on the preponderance of this evidence, I allow the impact

[4] It is important to note that most of the studies that comprise this literature offer no direct evidence on why "lower skilled" individuals are relatively more impacted, but tend to speculate that the observed relationship is due to lack of geographic mobility and/or because their employment is concentrated in sectors that are most impacted by economic changes.
[5] While not directly related, other studies which use national-level variation or focus on younger individuals tend to find similar demographic patterns (c.f., Blank, 1989; Acs and Wissoker, 1991; Freeman, 1991).
[6] Hoynes (2000) defines labor markets as MSAs and uses thirty-five MSAs in her analysis.

of local labor market conditions on various measures of health and health behaviors to vary across race and education groups, as discussed in Section III.

C. Related work

While the present work is related conceptually to the literature that investigates the impact of employment status on health, I limit my description to those studies that directly examine the connection between labor market conditions and health. In particular, I describe three recent studies, but defer comparisons of relevant findings to my own until Section V.

In the first rigorous study of its kind, Ruhm (2000) examines the impact of state-level unemployment rates on state-specific measures of total mortality and ten specific causes of death which account for roughly three-fourths of all deaths in the United States.[7] He finds evidence of a countercyclical relationship for total mortality and eight of the ten specific causes examined.[8] While automobile-related fatalities account for a substantial portion of the impact of changes in state unemployment rates on total mortality, the author finds that preventable causes of death account for an even greater portion of total deaths. Moreover, the author also examines age-specific death rates and finds that fatalities among those aged 20 to 44 are most sensitive to changes in state labor market conditions, consistent with the idea that his estimates are capturing a labor market phenomenon.

Second, Ruhm (2001), using data on individuals residing in thirty-one "large" MSAs from the 1972-1981 National Health Interview Surveys (NHIS) finds evidence of

[7] The author also explores behavioral reasons that might explain these findings, but the majority of this work is included in a separate study which I discuss in detail later in this section.
[8] The two exceptions are cancer and suicide. He finds that the suicide rate varies directly with state unemployment rates, suggesting that mental health is procyclical in nature.

a countercyclical relationship between state unemployment rates and several indicators of physical health including medical care utilization (e.g., hospital episodes and doctor visits), unhealthy days (e.g., restricted-activity days and bed days) and whether an individual experienced an acute, but not chronic, medical condition. The author allows the impact of state unemployment rates to vary across certain groups and find these relationships are most pronounced for males, employed persons and working-aged individuals.[9] Finally, consistent with his earlier finding regarding suicide, the author finds that non-psychotic mental disorders rise with increases in state unemployment rates and concludes that this represents "some evidence that mental health is procyclical."

Of the three studies described, Ruhm (2003) is most relevant to the present work because of the greater overlap in outcomes examined.[10] Using data on individuals aged eighteen and older from the 1987-2000 waves of the Behavioral Risk Factor Surveillance Survey (BRFSS), the author finds countercyclical relationships between state unemployment rates and several health behaviors. In particular, he finds systematic relationships for smoking, physical inactivity and weight related health.[11] Consistent with Ruhm (2001), he finds that these relationships are, generally speaking, most pronounced for males and employed persons.[12] Finally, the author presents evidence that suggests the impacts are considerably larger in the first half of the period in question. In

---

[9] Consistent with these findings, he reports systematic evidence that among chronic conditions "back disorders" are countercyclical.

[10] The exception is that Ruhm (2003) does not examine outcomes related to mental health.

[11] The statistical significance of reported estimates is not completely clear since the author reports standard errors that cluster on state of residence *and* month of interview, rather than state alone. In footnote 28 on page 10, the author notes that when he clustered on state of residence only, reported standard errors increased considerably. In particular, standard errors for smoking-related outcomes rose by 8 to 26 percent, by 33 to 50 percent for weight-related outcomes and by 165-186 percent for outcomes related to physical inactivity.

[12] However, the impact for individuals who are "not employed" can only be inferred since these estimates are not reported separately. In addition, it is difficult to judge whether, as in Ruhm (2001), working aged individuals are most impacted by fluctuations in state unemployment rates since the author does not report models by age group.

particular, estimates from models that include only observations for the years 1987 to 1994 are considerably larger in magnitude than estimates that include all years.[13] This is especially true in models that examine current smoking behavior and obesity.

II. Data

I use annual cross-sectional data from the National Health Interview Survey (NHIS) for the years 1997 to 2001, inclusive. While the NHIS dates back to 1972, it was redesigned in the middle 1990s, with 1997 the first wave following this revision. I use the adult sample which consists of annual surveys of thirty to thirty-five thousand individuals. To obtain a more localized measure of labor market conditions, I limit my analysis to individuals living in Level A or "large" metropolitan statistical areas (MSAs), for whom MSA of residence is publicly available.[14] This restriction yields between fifty and fifty-five percent of the overall NHIS sample, depending on the year in question. In the following paragraphs, I describe my key variables, focusing on measures of health that may fluctuate with changing labor market conditions and the MSA-level unemployment rate, which I use as a proxy for these conditions. Finally, I provide detailed information on my analysis sample.

A. Health measures

Conditional on availability, I focus on measures of health that may vary over short periods of time and whose diagnosis is independent of access to medical care. These measures can be grouped into three general categories: weight-related health, psychological well-being and self-reported health. I also examine a large set of health

---

[13] Estimates from models that include observations for the years 1995 to 2000 are not presented separately.
[14] Level A MSAs have at least one million residents. In 1997, they contained roughly 52 percent of the U.S. population.

behaviors which includes cigarette smoking, heavy alcohol consumption and frequency of physical exercise.

In terms of weight-related health, I focus on body mass index (BMI) and clinically-relevant thresholds based upon it. BMI is defined as the ratio of one's weight in kilograms to their height in meters squared. While BMI is preferred to body weight, and is a generally-accepted metric to assess weight-related health, it has certain shortcomings. First, BMI might not be a valid measure for some individuals, perhaps due to differences in body type or composition. If not, widely-used thresholds at the upper and lower tails of the distribution may misrepresent weight-related health. Second, BMI information in the NHIS is constructed from self-reports of height and weight, so it is subject to measurement error (Cawley, 1999).[15] In particular, it is likely that heavier individuals tend to under-report weight while lighter individuals over-report it. As noted by Lakdawalla and Philipson (2002), such systematic reporting may attenuate estimated coefficients rather than merely reduce their precision, as with classical measurement error in the dependent variable.

In addition to BMI, itself, I examine three thresholds of clinical interest, including underweight (BMI $\leq$18.5), overweight (BMI $\geq$ 25) and obesity (BMI $\geq$ 30). I also combine these thresholds to examine what happens to the fraction of individuals whose body weight falls in a "healthy" range. In particular, I model four overlapping ranges—BMI between 18.5 and 25, BMI between 18.5 and 30, BMI between 20 and 25 and BMI between 20 and 30. To the extent that local labor market conditions lead to weight gain

[15] While height and weight are self-reported, they were gathered via in-person interviews rather than, say, over the phone. It is likely that such interviews constrain individuals' ability to misreport their height and weight.

in some individuals and weight loss in others, these are useful measures of weight-related health.

With respect to psychological well-being, the NHIS includes six questions that assess an individual's state of mind in the month prior to being interviewed. These questions comprise the K6 Non-specific Psychological Distress scale which was designed to identify individuals who are likely to have both a diagnosable mental disorder and significant impairment. Validation studies show that this particular scale is at least as effective as more comprehensive and more established scales in diagnosing "serious mental illness" (Kessler et al., 2003).[16] The six questions that comprise the K6 scale are as follows. During the past 30 days, how often did you feel....

...so sad that nothing could cheer you up?

...hopeless?

...worthless?

...restless or fidgety?

...nervous?

...that everything was an effort?

Legitimate responses include "all of the time", "most of the time", "some of the time", "a little of the time" and "never". To assess how within-MSA changes in local unemployment rates affect reporting patterns, I parameterize responses to each of these six questions into three separate dichotomous indicators. In particular, I estimate three sets of models where the dependent variables are "most of the time, or more frequently", "some of the time, or more frequently", and "never".

---

[16] Kessler et al. (2003) provide evidence that the K6 scale is at least as effective in diagnosing "serious mental illness" as the longer K10 scale as well as the Composite International Diagnostic Interview Short-Form (CIDI-SF) and the World Health Organization Disability Assessment Schedule (WHO-DAS).

Measures of self-reported health are limited in the NHIS, but include a question that asks respondents to compare their current health to their health twelve months prior to being interviewed. The actual question asks, "Compared to twelve months ago, would you say your health is better, worse or about the same?" I categorize responses to this question as two dichotomous indicators—one which indicates if health has become "better" and one which indicates if health has become "worse". Since the question refers to the prior year, I use annual, rather than quarterly, measures of MSA-level unemployment rates.

Health behaviors analyzed include cigarette smoking, alcohol consumption and various measures of physical exercise. I label someone a smoker if he reports smoking cigarettes on at least some days per week. Since the cigarette excise tax rate has been shown to be an important determinant of smoking behavior, I also include state-level cigarette taxes in these models.[17] Detailed information on alcohol consumption is somewhat less available in the NHIS and I focus on measures that represent "heavy" drinking. In particular, I model the number of days in the twelve months prior to being interviewed that an individual consumed five or more alcoholic drinks. I also model two thresholds based on this measure—whether the individual has participated in any days of heavy drinking in the past year and whether he has engaged in fifty or more such days over the same time frame. The latter measure is intended to capture heavy drinking that occurs on a fairly regular basis. Finally, information on exercise includes "moderate" and "vigorous" exercise as well as information on strength training. Moderate exercise is

[17] More importantly, there is substantial variation in taxes over this period and this variation may be correlated with economic conditions (e.g., states raising sin taxes during an economic downturn). Note also that I use state population-weighted averages to assign cigarette tax rates to individuals residing in the fourteen MSAs that overlap one or more states.

defined as exercise that causes "only light sweating or slight to moderate increases in breathing or heart rate" while vigorous exercise is defined as exercise that causes "heavy sweating or large increases in breathing or heart rate". Data on "moderate" and "vigorous" exercise include information on the number of times per week an individual engages in either type of activity for at least twenty minutes.[18] Data on strength training include no time component and refer only to the number of times per week an individual engages in such activity, irrespective of the time spent at each session. For completeness, I define three dependent variables that measure the frequency of each of these three types of exercise: Any times per week, 3 or more times per week, and 5 or more times per week.

B. Local labor market conditions

I use MSA-level unemployment rates from the Bureau of Labor Statistics' Local Area Unemployment Statistics database as a proxy for local labor market conditions. As indicated earlier, previous work relating the unemployment rate to health has focused on state-level measures, implicitly treating the state as the labor market of relevance. As seen in Figure 1, the unemployment rate falls and then rises, for a nearly U-shaped relationship over the period in question. Examining this pattern by groups defined by predicted employment status (Figure 2), race (Figure 3) and level of education (Figure 4) shows that the same U-shaped relationship obtains. The roughly parallel lines indicate very similar experiences over time, though the gaps indicate level differences in average unemployment regime.

[18] For the first six months of 1997 questions regarding moderate and vigorous exercise were asked in terms of "at least ten minutes" per day, rather than the twenty minutes asked in all subsequent survey periods. For consistency, I drop individuals interviewed in the first six months of 1997 from these models.

C. Analysis sample

Restricting my sample to those men who live in large MSAs, as described above, yields 38,101 men from five years of data. I further limit my sample to individuals between twenty-four and fifty-nine years old. On the upper end of this range, I aim to avoid retirement issues which may be affected by local labor market conditions. On the lower end, I want to avoid schooling or training issues, since labor market conditions may also influence these decisions. These age restrictions reduce my sample to 27,159 men. Since I include indicator variables for missing data on other covariates, this figure represents the sample I use to generate most estimates discussed in Section IV, though note that missingness in the dependent variable, itself, reduces sample size in specific models.

III. Empirical strategy

Unobserved heterogeneity is a primary concern in relating local labor market conditions and health. More precisely, the concern is that unobserved labor market characteristics that are correlated with the unemployment rate and exert an independent influence on health will result in biased estimates. For example, some areas may experience both poor health and high unemployment though no causal relationship exists. In a single cross-section of data, this would induce a procyclical relationship where none may exist. The repeated cross-sectional nature of NHIS data allows for inclusion of MSA fixed effects, which will eliminate the troublesome heterogeneity if it is time invariant over the five years in question.

With this in mind, a model that bases statistical identification on within-MSA variation in the unemployment rate is given by:

$$\mathbf{H}_{ijqt} = \boldsymbol{\tau}\mathbf{U}_{jqt} + \boldsymbol{\beta}\mathbf{X}_{ijqt} + \boldsymbol{\mu}_j + \boldsymbol{\theta}_{qt} + \boldsymbol{\varepsilon}_{ijqt} \qquad (1)$$

Here, i indexes the individual, j MSA of residence, q quarter surveyed, and t year surveyed. **H** represents the relevant measure of health or health behavior, **U** the MSA-specific unemployment rate, **X** a set of individual and MSA-specific covariates, $\boldsymbol{\mu}$ is a vector of MSA fixed effects, $\boldsymbol{\theta}$ year-specific quarter fixed effects and ε captures unobserved determinants of health.[19] With the exception of models that condition on its elements, the vector **X** includes controls for age, race, education level, prior year's household income as a fraction of the poverty line, marital status and employment status. All models include MSA and time fixed effects as specified in equation (1) and estimation of all models uses the cluster option of Stata with clusters defined as entire MSAs, rather than MSA-year cells.

This specification, however, has one prominent drawback. It imposes the same relationship between local labor market conditions and health for all individuals. As discussed earlier, previous work on the distributional impacts of economic conditions on employment-related outcomes suggests that lower skilled individuals, particularly non-white and less educated individuals, are most affected by such fluctuations. To address this shortcoming, I take two distinct approaches. First, I allow the effect of the local unemployment rate to vary by an individual's "exposure" to labor market fluctuations. Since exposure is not directly observable, I measure it via an individual's predicted employment status. In particular, I first estimate a cross-sectional model of employment status using 1997 data, which is intended to capture the data generating process for employment status prior to subsequent fluctuations in the local unemployment rate.

[19] Aside from differences in the geographic definition of a labor market (MSA versus entire state), this specification is conceptually identical to that used by Ruhm (2003).

Using estimated coefficients from this model, I compute predicted employment probabilities for all individuals with useable employment and MSA of residence information.[20] Next, I split this distribution into deciles and estimate equation (1) separately for each of these ten groups.[21] As seen in Table A1, there are substantial demographic differences across the three groups listed. In particular, those least likely to be employed (i.e., 10th percentile) are more likely to be non-white, less educated and unmarried relative to individuals in the other two groups. Second, I estimate equation (1) separately by race and education level. Race groups include African American, Hispanic and white. Since individuals with education beyond a high school diploma, but less than a bachelor's degree are more like high school graduates in relevant health characteristics and health behaviors, I assign individuals to two educational groups—those with less than a bachelor's degree and those with a bachelor's degree or higher levels of education.[22]

IV. Results

Relevant empirical evidence suggests the employment status of certain groups is relatively more affected by economic fluctuations. Consistent with this general finding, I allow the estimated effect of local labor market conditions to vary across groups defined

---

[20] More precisely, models that generate the predicted probabilities are linear probability models and the predicted probability is given generally by $X_{ijqt}'\beta_{97}$, where β_{97} is the vector of coefficient estimates from the cross-sectional model and X_{ijqt} represents the plausibly exogenous characteristics (i.e., age, race, education level, marital status) of individual i residing in MSA j in quarter q of year t. To address the possibility that estimated relationships between employment status and other covariates may change with fluctuations in the unemployment rate over time, I also pool all years of data and estimate these probabilities with models that include time fixed effects only. The correlation between these two sets of probabilities exceeds 0.99.
[21] Since none of the eight deciles between the first (lowest) and the tenth (highest) show any systematic relationship between local labor market conditions and any measure of health or any health behavior, I combine individuals in these deciles into one group in relevant tables.
[22] For example, while twenty-two percent of respondents with more than a high school diploma, but less than a bachelor's degree and twenty-three percent of high school graduates report being clinically obese, only sixteen percent of those with a bachelor's degree report likewise. Corresponding figures for cigarette smoking are, respectively, twenty-eight, twenty-three and twelve percent.

by their predicted employment status, race and educational attainment. I report estimates in a similar fashion; this section contains three sub-sections, each corresponding to one of these three delineations. In each sub-section, I present results on the impact of the local unemployment rate on weight-related health, psychological well-being, self-reported health as well as an extensive set of health behaviors. I interpret estimates in the context of a one percentage point increase in the local unemployment rate. Given the large volume of estimates, I limit my discussion to the estimated impacts of local labor market conditions.

A. Estimates by predicted employment status

Tables 1A-1D are organized as follows: Column 1 represents individuals in the lowest decile of the predicted employment distribution (i.e., those least likely to be employed), Column 2 represents individuals in the highest decile of this distribution (i.e., those most likely to be employed) and Column 3 represents individuals who fall in the eight deciles between these two extremes.[23] I collapse these eight deciles into one group to facilitate the discussion of estimates and also because I detect no systematic patterns in any of the individual deciles for any outcome.

Table 1A displays the estimated effect of local labor market conditions on weight-related health. The estimates imply that those least likely to be employed experience increases in weight when the local unemployment rate rises, though I find no evidence of a systematic relationship for the other two groups. For individuals in the lowest decile, I estimate that a one percentage point increase in the local unemployment rate leads to a 1.34 pound gain, on average. While the relevant coefficient is statistically insignificant,

[23] Table 1B is an exception as it contains three columns for each of these three groupings for a total of nine columns. I describe its structure below.

note that it reflects an average effect of local labor market conditions on body weight. If, for example, a rising unemployment rate leads to weight gains for some individuals and losses for others, this average effect will be attenuated. Moreover, focusing on BMI masks where in the weight distribution prospective gains or losses may be occurring.[24] As a result, it is more informative to examine clinically-relevant thresholds based on BMI. I model three such thresholds: clinical measures of underweight (BMI≤18.5), overweight (BMI≥25) and obesity (BMI≥30). Estimates from the first column of Table 1A suggest increases in the local unemployment rate do indeed have dual impacts on weight for this group. In particular, these estimates suggest not only increases in the fraction overweight and obese, but also increases in the fraction underweight, though the latter estimate is only marginally significant at conventional levels. Focusing on the overweight and obese thresholds, relevant coefficients imply percentage point increases of 3.5 and 2.1, respectively. In percentage terms, these represent increases of roughly six and nine percent. As described earlier, I also define four ranges of BMI that represent "healthy" body weights. Corresponding estimates are all negative and statistically significant at conventional levels. Focusing on the broadest of these ranges (BMI between 18.5 and 30), I find that a one percentage point increase in the local unemployment rate leads to roughly a 2.7 percentage point decrease in the fraction of those least likely to be employed in this range. In percentage terms, this represents a decrease of about four percent.

Table 1B shows the impact of local labor market conditions on various measures of psychological well-being. The table is constructed as follows: The first six rows

[24] For example, increases in body weight among those clinically underweight likely represent different changes in weight-related health than similar gains among obese or near obese individuals.

represent the questions that comprise the K6 Scale of Non-specific Psychological Distress, which was described in Section II. The final row collapses responses to each of these six questions into a single metric which represents whether an individual reports *any* of the six indicated emotions at the frequencies indicated. For each question, I define three different dependent variables based on possible responses. These dependent variables correspond to columns labeled "most", "some" and "never".[25] So, for each of the three groups defined by their predicted employment status, I estimate twenty-one separate models. Table 1B displays estimated coefficients on local unemployment rate from each of these models.[26]

Estimates from Table 1B exhibit a consistent sign pattern. With a single exception, coefficient estimates in the "most" and "some" models are positive, while corresponding estimates in the "never" models are always negative. This pattern indicates that all three groups experience diminished psychological well-being when the local unemployment rate increases. Closer inspection, however, shows that this pattern is most pronounced for those in the lowest predicted employment decile, where all seven coefficients in the "some" models are statistically different from zero at conventional levels.[27] Beyond statistical significance, estimates for this group are also larger in magnitude than those of the other two groups. Focusing on the "some" models, relevant estimates imply a one percentage point increase in the local unemployment rate leads to 3.4, 3.3, 2.5, 3.5, 3.5 and 3.8 percentage point increases in the fraction responding

---

[25] Respectively, these represent the following frequencies: "most of the time, or more frequently", "some of the time, or more frequently", and "never".

[26] To be clear, the first column of Table 1B represents the estimated impact of local unemployment rate on whether an individual reports being "so sad that nothing could cheer him up" in the past thirty days "most of the time, or more frequently".

[27] In addition, three of the seven coefficients in the "most" and "never" models are statistically different from zero.

affirmatively in models of sadness, hopelessness, worthlessness, restlessness, nervousness, and feelings of effort, respectively. In percentage terms, these represent increases of fifteen, twenty-four, twenty-two, fifteen, sixteen and seventeen percent, respectively.

Table 1C presents estimates of the impact of local labor market conditions on self-reported health. Due to data availability, I analyze two measures of self-reported health—whether one reports their health getting "better" in the twelve months prior to being interviewed and whether one reports their health getting "worse" over the same period. Since the relevant question refers to the prior year, I use annual, rather than quarterly, variation in the local unemployment rate. Relevant estimates imply that those least likely to be employed are more likely to report worsened health and less likely to report improved health as the local unemployment rate increases. Estimates for those in the combined eight deciles exhibit a similar pattern, while estimates for those most likely to be employed are mixed in sign. Given their lack of precision, however, these estimates provide very limited evidence of a systematic relationship between self-reported health and local labor market conditions.

Table 1D displays results related to available health behaviors. Overall, there is little evidence that the local unemployment rate impacts any of these behaviors for any of the three groups. This is especially evident for the three measures of physical exercise—moderate exercise, vigorous exercise and strength training—and their various levels of intensity—any times per week, three or more times per week and five or more times per week. One notable exception, however, is smoking behavior. My estimates imply an increase in smoking behavior for those least likely to be employed, but reductions in smoking for those in the highest employment decile. In particular, a one percentage point increase in the local unemployment rate is associated with roughly a 2.7 percentage point increase for those in the lowest employment decile and a 2.3 percentage point reduction for those most likely to be employed. Respectively, these figures represent an eight percent increase and a twenty-one percent decrease.

B. Estimates by race

In the following paragraphs, I describe findings from models estimated separately for African American, Hispanic and white men. In general, the estimates imply gains in body weight and reduced psychological well-being for African American males in response to worsening labor market conditions, but suggest no such evidence for their Hispanic and white counterparts. In addition, there is little evidence of a systematic relationship between the local unemployment rate and self-reported health or any health behaviors for any of the three groups.

Table 2A presents the estimated effect of local labor market conditions on weight-related health. The general pattern of estimates implies African American men gain weight when the local unemployment rate rises, but suggest no systematic relationship for either Hispanic or white men. As seen in Table 2A, the relevant estimate in the log BMI model (row 1) implies an average weight gain of 1.8 pounds in response to a one percentage point increase in local unemployment rate. In terms of distributional impacts, this average gain appears to be generated towards the right tail of the BMI distribution as a one percentage point increase in the local unemployment rate leads to 1.9 and 2.2 percent gains in the fraction of African American males who are clinically overweight and obese, respectively. In percentage terms, these gains are about three and nine percent. Finally, all four models of "healthy" body weight imply reductions in the fraction of African American males in these ranges. For example, a one percentage point increase in the local unemployment rate is associated with nearly a 2.2 percentage point decrease in the fraction with BMI between 18.5 and 30, a decrease of nearly three percent.

Table 2B displays the estimated impact of local labor market conditions on psychological well-being. The general pattern of estimates suggests that all three groups experience reduced psychological well-being when the local unemployment rate rises. The pattern, however, is most pronounced for African American and white men. With respect to African Americans, all coefficient estimates in the "most" and "some" models are positive, while all corresponding estimates in the "never" models are negative. Focusing on the "most" models, four of the seven models exhibit estimates that are statistically different from zero at conventional levels. The last row of the first column of Table 2B implies that the fraction of African American males who report any of the relevant six feelings "most of the time, or more frequently" increases by nearly 1.8 percentage points in response to a one percentage point increase in the local unemployment rate, which represents an increase of roughly twenty-two percent. For white men, the most consistent evidence of reduced psychological well-being is seen in the "some" models. Here, five of the seven models yield coefficient estimates that are statistically significant. While this seems inconsistent with the idea that groups whose employment is most impacted by local labor market fluctuations should have their health most impacted, further examination reveals that estimates in Table 2B are driven by white men with relatively low educational attainment. In particular, when I divide white males into two groups according to their education, I find evidence of reduced psychological well-being among those with less than a bachelor's degree, but weaker evidence among those with at least this amount of formal education.[28] For example, while the estimated coefficient in the "some" model where the dependent variable reflects an affirmative response to at least one of the six indicated emotions is 0.0140 (t=2.54), corresponding estimates for low and high education groups are 0.0168 (t=2.26) and 0.0113 (t=1.21), respectively.

Table 2C displays estimates regarding self-reported health. While estimates indicate a procyclical relationship for African American and Hispanic men, the relevant estimates are not precisely estimated. Estimates for white males imply the fraction of those reporting worse health increases with increases in the local unemployment rate. However, as with measures of psychological well-being, this result is driven by white men with lower educational attainment. In particular, the coefficient estimate for these men is 0.0132 (t=2.79), while the corresponding estimate for white men with at least a bachelor's degree is 0.0002 (t=0.05). So, if there is any evidence of poorer self-reported health among white men, it is confined to those with less than a bachelor's degree.

Table 2D presents estimates related to various health behaviors. Again, there is almost no evidence of a systematic relationship between these health behaviors and local labor market conditions. The lone exception is smoking behavior among African American males. The relevant estimate suggests that their smoking increases in response to increases in the local unemployment rate. In particular, a one percentage point increase in the local unemployment rate is associated with a nearly thirteen percent gain in the fraction of African American males who smoke.

---

[28] In the next subsection, I estimate all models by education level and discuss relevant issues.

C. Estimates by education level

In this subsection, I describe findings from models estimated separately by educational attainment. As discussed in section II, I define two groups—those with less than a bachelor's degree and those with at least a bachelor's degree—to represent "high" and "low" education groups. While there is some evidence that more educated males gain weight when labor market conditions worsen, these gains seem to be generated by African American males with at least a bachelor's degree. More consistent evidence suggests reduced psychological well-being among less educated males. Once again, there is little evidence of a systematic relationship between the local unemployment rate and the self-reported health or health behaviors of either of these two groups.

Table 3A displays estimates of the impact of local labor market conditions on weight-related health. The general pattern of estimates suggests that relatively more educated men gain weight when the local unemployment rate rises, while there is no such evidence for men with less than a bachelor's degree. The relevant coefficient estimate in Table 3A implies an average weight gain of about one pound when the local unemployment rate increases by one percentage point. Note, however, that the estimated effect appears to be driven by African American men in the high education group. When these models are estimated separately by race, the corresponding coefficients are 0.0159 (t=1.94), 0.0046 (t=0.26) and 0.0039 (t=1.32) for highly educated African American, Hispanic and white men in the log BMI specification. Estimates for other measures of weight-related health show a similar pattern, though none are estimated precisely.

Table 3B shows the estimated effect of local labor market conditions on psychological well-being. A consistent sign pattern is evident as coefficient estimates in all but two of the "most" and "some" models are positive, while all corresponding estimates in the "never" models are negative. This pattern suggests that both education groups experience reduced psychological well-being when the local unemployment rate rises. Note, however, that the pattern is more pronounced for individuals with less than a bachelor's degree, where five of the seven "some" models yield estimates statistically different from zero at conventional levels.[29] Beyond precision, estimated coefficients for less educated men are consistently larger than those of their more educated counterparts. Focusing on the "some" models, relevant estimates imply a one percentage point increase in the local unemployment rate results in 0.7, 1.3, 0.8, 0.9, 1.4 and 1.1 percentage point gains in the fraction reporting feelings of sadness, hopelessness, worthlessness, restlessness, nervousness and effort, respectively, "some of the time, or more frequently". In percentage terms, these represent increases of about six, twenty-one, sixteen, five, ten and eight percent, respectively.

Table 3C presents estimates related to self-reported health. While the sign pattern of the relevant coefficients for less educated males indicates a procyclical relationship, they are not estimated very precisely. Estimates for more educated males are mixed as their signs imply increases in both the fraction reporting better and worse health when the local unemployment rate increases, though these estimates also are not precisely estimated.

Finally, Table 3D shows estimates related to health behaviors by education group. Consistent with earlier estimates, there is no systematic evidence of a relationship

[29] In addition, three of the seven coefficients in the "most" and two of the seven coefficients in the "never" models are statistically significant.

between the set of health behaviors examined and the local unemployment rate. In general, this is a consistent finding across all groups examined in this paper.

V. Discussion

As discussed in the previous section, I find systematic evidence of procyclical relationships for weight-related health and psychological well-being for a sample of men residing in the fifty-eight largest MSAs in the U.S. The relationships are most pronounced for those individuals in the lowest predicted employment decile and African Americans. In what follows, I discuss my most notable findings, how they relate to the two primary explanations for cyclicality in health and compare them to existing evidence.

In general, economists conceptualize body weight in terms of a simple production function where changes in weight are determined by the difference between the intake and expenditure of calories (c.f., Philipson and Posner, 1999). As noted, I find evidence of body weight increases in response to rising unemployment rates for those least likely to be employed and African Americans. Coupled with my finding of no systematic link between local unemployment rates and physical exercise, increased caloric intake is a natural implication. While this is plausible, perhaps through increased consumption of lower quality foods, the nature of the production function provides an alternative explanation. In particular, it is well-established that under extreme acute stress (e.g., if wounded or undergoing surgery) the human body stores calories more efficiently because of a recognized need for energy at a later point in time. More recent evidence suggests that this might also apply to prolonged episodes of less severe stress. If so, labor market related stress may affect weight independent of calories consumed or expended. In terms of their relation to the two primary explanations for cyclicality in health, my estimates appear to reject the time use hypothesis which, generally speaking, argues that the production of weight-related health should increase when labor market conditions worsen. To the extent that stress, rather than additional caloric intake, explains the observed increases, the estimates provide some support for the economic stress hypothesis. However, since I can not pinpoint the reason(s) for the observed weight gains with my data, it is impossible to distinguish between these two general explanations.

My finding that psychological well-being is procyclical in nature is seen more broadly, but is most pronounced for groups whose employment prospects have been found to be most impacted by labor market fluctuations. In particular, I find consistent evidence of reduced psychological well-being for those least likely to be employed, African Americans and individuals with less than a bachelor's degree. Relative to my weight-related health findings, this evidence appears more consistent with the economic stress hypothesis since time is likely not as prominent an input in the production of mental, as opposed to physical, health. In essence, the associated models are more direct tests of the economic stress hypothesis, as it relates to mental health.

Of my two major findings, one is consistent with existing evidence and the other contradicts it. My finding of reduced psychological well-being is consistent with Ruhm's (2000) finding that death by suicide is procyclical in nature as well as Ruhm's (2001) finding that non-psychotic mental disorders rise with worsening labor market conditions. Conversely, my finding of increased body weight is at odds with Ruhm (2003) which finds evidence of systematic reductions in weight in response to rising state unemployment rates. It is also inconsistent with his finding that physical exercise increases when labor market conditions deteriorate.

While it is difficult to identify reasons for these differences, I offer three possible explanations. First, I use a more localized measure of labor market conditions than existing work. In particular, I proxy these conditions with MSA, rather than state, unemployment rates. To the extent that there is variation in MSA unemployment rates that is not explained by corresponding state-level variation, it is possible my geographically narrower definition of a labor market accounts for any observed differences.[30] Second, my most consistent findings are for groups whose employment and employment prospects are most impacted by labor market fluctuations. One explanation why these groups are more sensitive to changing conditions is that they lack geographic mobility in response to such fluctuations. To the extent that this is true, MSA, rather than state, unemployment rates may provide a more appropriate characterization of the labor market conditions faced by individuals in these groups. Finally, structural change in the relationships examined may account for any observed differences. Ruhm (2003) provides some indirect evidence for this possibility. As noted, his finding that obesity is countercyclical appears to be considerably larger in magnitude for the period 1987-1994 than for the period 1995-2000, which more closely overlaps my period of interest.[31] As a result, observed differences may not be as large as they might otherwise appear.

---

[30] Relatedly, using state unemployment rates to proxy labor market conditions may be more useful in small, rather than large, states, to the extent that larger states contain more labor markets. Of the fifty-eight MSAs represented in my data, twenty-two are located in California, Florida, New York and Texas.
[31] Ruhm (2003) reports estimates for all sample years (1987-2000) and an earlier subset of these years (1987-1994) in his Table 3. Since he does not report estimates for 1995-2000 separately, I must infer these differences.

VI. Conclusions

In this paper, I present systematic evidence of procyclical relationships for weight-related health and psychological well-being for a sample of men living in the fifty-eight largest MSAs in the U.S. I find these relationships are most pronounced for groups previously found to be most affected by changing labor market conditions. In particular, my evidence is most consistent for those least likely to be employed and African Americans. As discussed, my findings have similarities to and differences with existing work, which suggests that physical health is countercyclical in nature while mental health is procyclical. Given the consistency of the findings regarding mental health, a deeper understanding of the long-run implications of changes in mental health is appropriate. For example, how long does the detrimental effect of worsening labor market conditions reduce mental health? Moreover, do such changes in mental health have implications for longer-run physical health? To the extent possible, such extensions should be examined in future related work.

Figure 1. Unemployment rate, 1997 to 2001

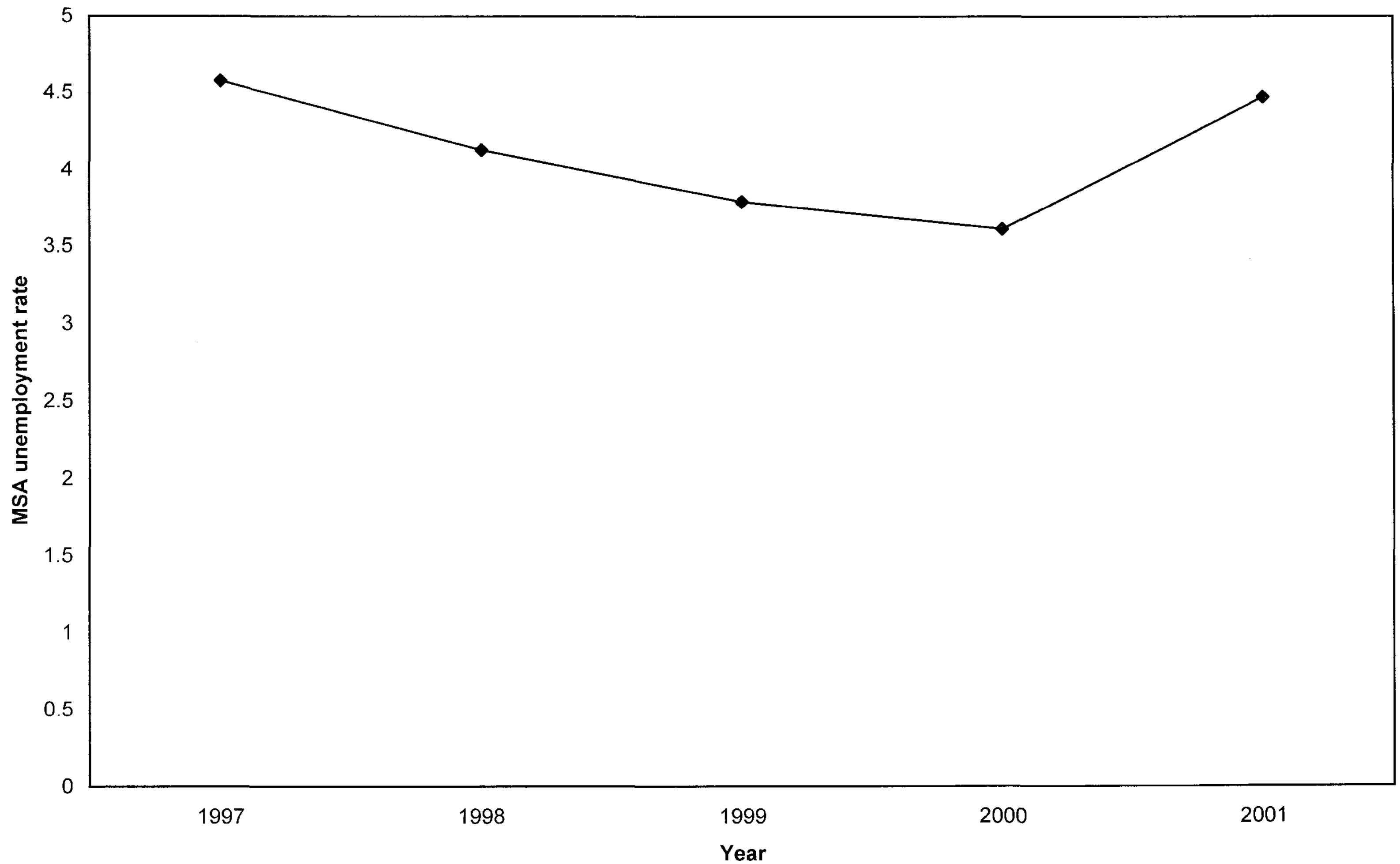

Figure 2. Unemployment rate by percentiles of the predicted employment distribution

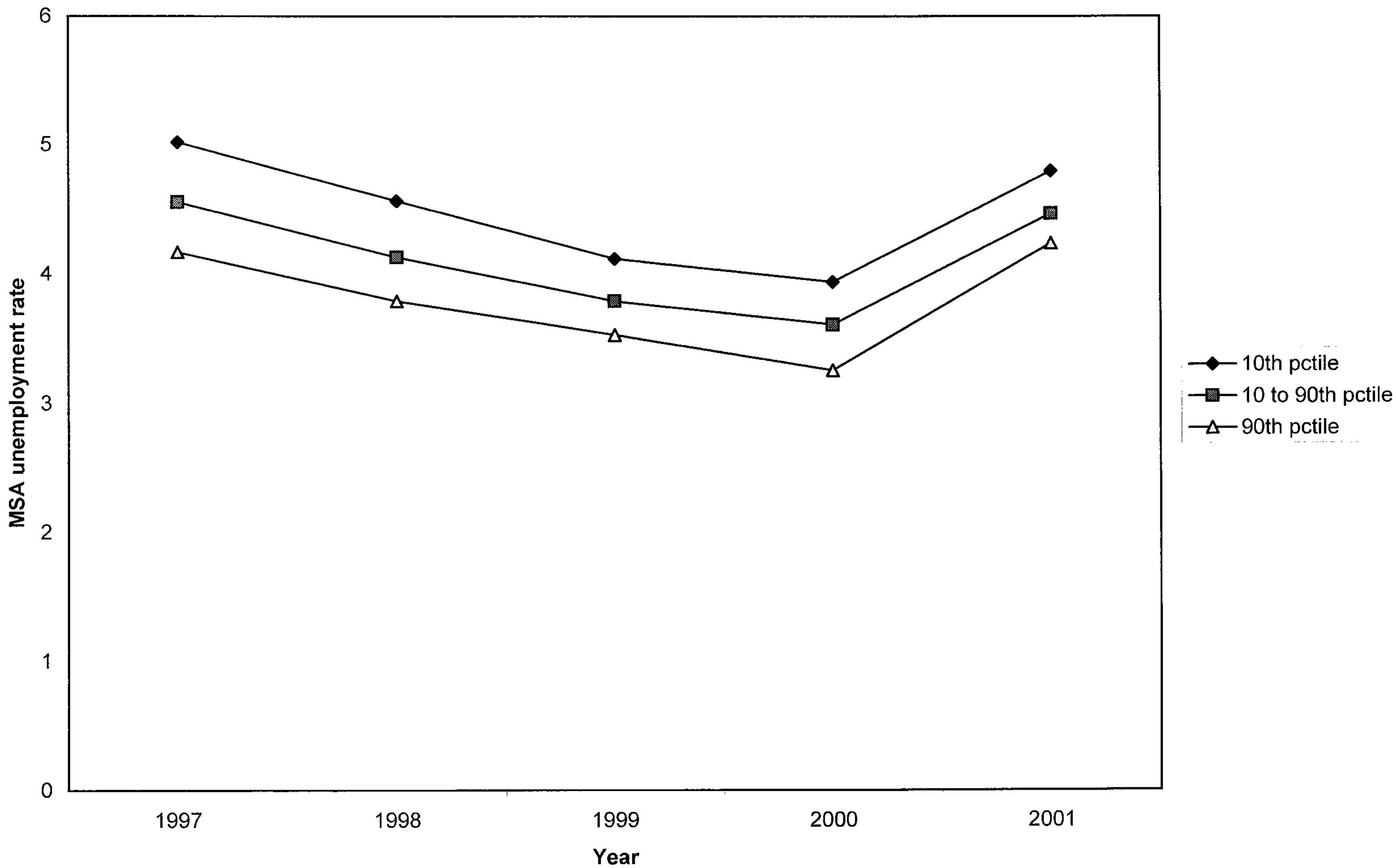

Figure 3. Unemployment rate by race

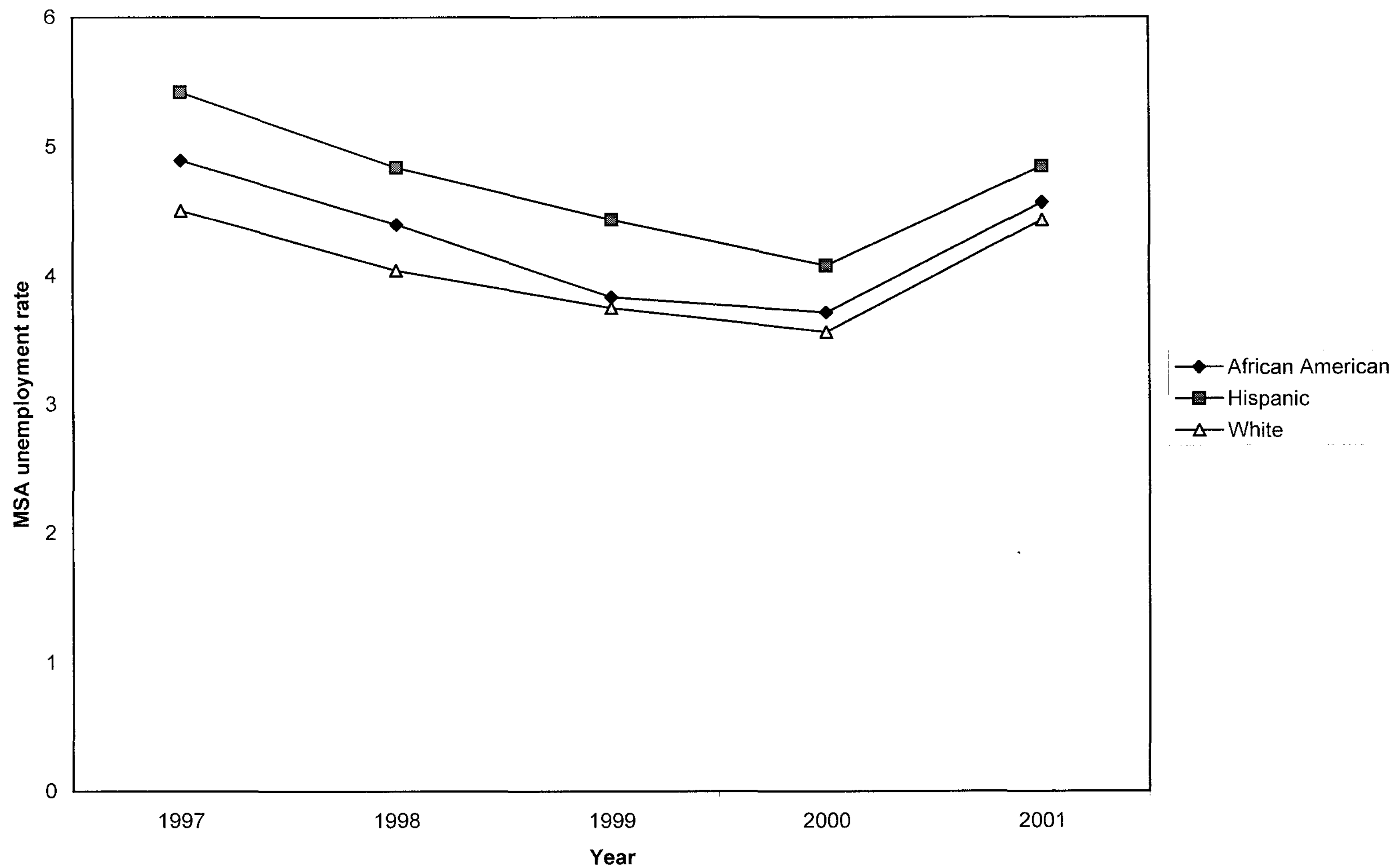

Figure 4. Unemployment rate by education level

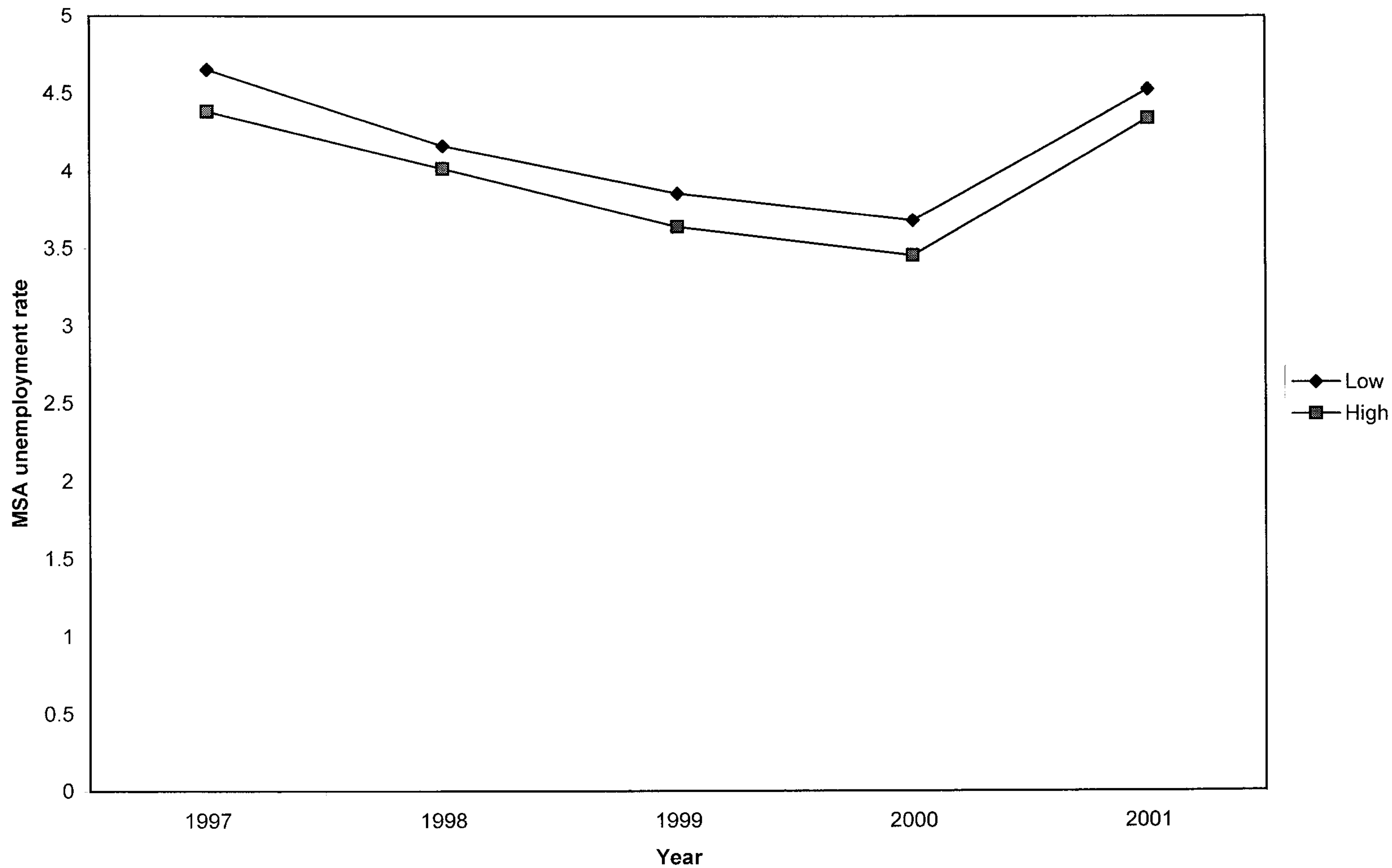

Table 1A. Estimated effect of MSA unemployment rate on weight-related health, by percentile of predicted employment distribution.

	10^{th} percentile	10^{th}-90^{th} percentile	90^{th} percentile
Log of BMI	0.0067	0.0012	-0.0026
	(0.0048)	(0.0016)	(0.0058)
BMI	0.1992	0.0408	-0.0480
	(0.1409)	(0.0487)	(0.1652)
BMI between 18.5 and 25	-0.0405	-0.0002	0.0100
	(0.0158)	(0.0059)	(0.0184)
BMI between 20 and 25	-0.0437	0.0005	0.0029
	(0.0160)	(0.0059)	(0.0176)
BMI between 18.5 and 30	-0.0268	-0.0020	-0.0010
	(0.0093)	(0.0051)	(0.0143)
BMI between 20 and 30	-0.0299	-0.0013	-0.0081
	(0.0095)	(0.0054)	(0.0128)
Underweight (BMI≤18.5)	0.0054	-0.0004	-0.0012
	(0.0032)	(0.0006)	(0.0029)
Overweight (BMI≥25)	0.0351	0.0006	-0.0088
	(0.0153)	(0.0057)	(0.0184)
Obese (BMI≥30)	0.0214	0.0024	0.0022
	(0.0090)	(0.0050)	(0.0145)
Sample size	2,648	21,129	2,685

Notes: All estimates are from OLS regressions. Models include controls for education, income relative to poverty line, age, race, marital status, employment status, and fixed effects for MSA and year-specific quarter of interview. Samples include males aged 24-59. Standard errors adjusted for non-independence of observations within MSAs are in parentheses.

Table 1B. Estimated effect of MSA unemployment rate on psychological well-being, by percentile of predicted employment distribution.

	10th percentile			10th-90th percentiles			90th percentile		
	Most	Some	Never	Most	Some	Never	Most	Some	Never
Sad	0.0152	0.0340	-0.0234	0.0010	0.0027	-0.0006	0.0003	0.0002	-0.0035
	(0.0067)	(0.0141)	(0.0192)	(0.0023)	(0.0042)	(0.0070)	(0.0026)	(0.0074)	(0.0098)
Hopeless	0.0105	0.0333	-0.0371	0.0024	0.0059	-0.0052	0.0019	0.0055	-0.0086
	(0.0045)	(0.0084)	(0.0120)	(0.0018)	(0.0031)	(0.0029)	(0.0022)	(0.0050)	(0.0077)
Worthless	0.0085	0.0246	-0.0274	0.0007	0.0035	-0.0005	0.0006	0.0024	-0.0063
	(0.0052)	(0.0095)	(0.0112)	(0.0019)	(0.0021)	(0.0029)	(0.0017)	(0.0043)	(0.0068)
Restless	0.0039	0.0350	-0.0271	0.0019	0.0026	0.0010	0.0041	0.0156	-0.0521
	(0.0081)	(0.0149)	(0.0200)	(0.0027)	(0.0039)	(0.0079)	(0.0069)	(0.0134)	(0.0143)
Nervous	0.0123	0.0348	-0.0402	0.0012	0.0048	-0.0071	-0.0036	0.0045	-0.0199
	(0.0072)	(0.0106)	(0.0158)	(0.0022)	(0.0035)	(0.0058)	(0.0042)	(0.0126)	(0.0186)
Effort	0.0131	0.0376	-0.0236	0.0042	0.0040	-0.0047	0.0024	0.0069	-0.0361
	(0.0088)	(0.0135)	(0.0171)	(0.0024)	(0.0037)	(0.0075)	(0.0055)	(0.0110)	(0.0143)
Any of the above	0.0250	0.0401	-0.0181	0.0061	0.0088	-0.0052	0.0031	0.0248	-0.0350
	(0.0112)	(0.0159)	(0.0218)	(0.0033)	(0.0052)	(0.0106)	(0.0090)	(0.0156)	(0.0145)

Notes: All estimates are from OLS regressions. Models include controls for education, income relative to poverty line, age, race, marital status, employment status, and fixed effects for MSA and year-specific quarter of interview. Samples include males aged 24-59. Standard errors adjusted for non-independence of observations within MSAs are in parentheses.

Table 1C. Estimated effect of MSA unemployment rate on self-reported health, by percentile of predicted employment distribution.

	10^{th} percentile	10^{th}-90^{th} percentiles	90^{th} percentile
Health "worse"	0.0131	0.0017	0.0069
	(0.0095)	(0.0027)	(0.0083)
Health "better"	-0.0028	-0.0040	0.0087
	(0.0130)	(0.0048)	(0.0113)
Sample size	2,695	21,623	2,716

Notes: All estimates are from OLS regressions. MSA unemployment rate is measured annually, rather than quarterly, to match the phrasing of the relevant question. Models include controls for education, income relative to poverty line, age, race, marital status, employment status, and fixed effects for MSA and year-specific quarter of interview. Samples include males aged 24-59. Standard errors adjusted for non-independence of observations within MSAs are in parentheses.

Table 1D. Estimated effect of MSA unemployment rate on selected health behaviors, by percentile of predicted employment distribution.

	10th percentile	10th-90th percentiles	90th percentile
Current smoker	0.0273	0.0053	-0.0231
	(0.0161)	(0.0047)	(0.0091)
Any days with 5+ drinks	0.0058	-0.0064	0.0363
	(0.0133)	(0.0077)	(0.0177)
Days with 5+ drinks	-1.605	-0.1217	-1.733
	(1.7750)	(0.5604)	(1.3703)
Moderate exercise, any times per week	0.0068	0.0025	-0.0038
	(0.0131)	(0.0081)	(0.0162)
Moderate exercise, 3+ times per week	0.0049	-0.0024	-0.0069
	(0.0134)	(0.0077)	(0.0157)
Moderate exercise, 5+ times per week	0.0035	0.0009	0.0013
	(0.0114)	(0.0059)	(0.0164)
Vigorous exercise, any times per week	-0.0089	-0.0042	0.0083
	(0.0177)	(0.0113)	(0.0172)
Vigorous exercise, 3+ times per week	0.0010	-0.0058	0.0143
	(0.0162)	(0.0084)	(0.0183)
Vigorous exercise, 5+ times per week	-0.0008	-0.0038	0.0045
	(0.0096)	(0.0053)	(0.0174)
Strength training, any times per week	-0.0087	0.0066	-0.0066
	(0.0191)	(0.0051)	(0.0155)
Strength training, 3+ times per week	-0.0067	-0.0042	-0.0012
	(0.0119)	(0.0048)	(0.0166)
Strength training, 5+ times per week	-0.0015	-0.0002	0.0102
	(0.0072)	(0.0035)	(0.0111)

Notes: All estimates are from OLS regressions. Models include controls for education, income relative to poverty line, age, race, marital status, employment status, and fixed effects for MSA and year-specific quarter of interview. Smoking equations also include state-level excise tax on cigarettes as a covariate. Samples include males aged 24-59. Standard errors adjusted for non-independence of observations within MSAs.

Table 2A. Estimated effect of MSA unemployment rate on weight-related health, by race.

	African-American	Hispanic	White
Log of BMI	0.0085	-0.0023	0.0018
	(0.0035)	(0.0038)	(0.0018)
BMI	0.2699	-0.0476	0.0504
	(0.1027)	(0.1117)	(0.0526)
BMI between 18.5 and 25	-0.0190	-0.0062	-0.0023
	(0.0102)	(0.0137)	(0.0078)
BMI between 20 and 25	-0.0151	-0.0069	-0.0033
	(0.0102)	(0.0126)	(0.0073)
BMI between 18.5 and 30	-0.0216	0.0082	-0.0054
	(0.0092)	(0.0124)	(0.0056)
BMI between 20 and 30	-0.0178	0.0075	-0.0064
	(0.0092)	(0.0108)	(0.0056)
Underweight (BMI≤18.5)	-0.0003	0.0020	0.0001
	(0.0017)	(0.0011)	(0.0010)
Overweight (BMI≥25)	0.0192	0.0042	0.0022
	(0.0100)	(0.0132)	(0.0073)
Obese (BMI≥30)	0.0219	-0.0102	0.0054
	(0.0091)	(0.0116)	(0.0052)
Sample size	3,998	4,091	15,810

Notes: All estimates are from OLS regressions. Models include controls for education, income relative to poverty line, age, race, marital status, employment status, and fixed effects for MSA and year-specific quarter of interview. Samples include males aged 24-59. Standard errors adjusted for non-independence of observations within MSAs are in parentheses.

Table 2B. Estimated effect of MSA unemployment rate on psychological well-being, by race.

	African-American			Hispanic			White		
	Most	Some	Never	Most	Some	Never	Most	Some	Never
Sad	0.0066	0.0059	-0.0201	0.0051	0.0094	0.0103	0.0017	0.0062	-0.0058
	(0.0040)	(0.0080)	(0.0126)	(0.0054)	(0.0085)	(0.0160)	(0.0022)	(0.0043)	(0.0062)
Hopeless	0.0106	0.0077	-0.0116	0.0040	0.0081	-0.0050	0.0013	0.0090	-0.0115
	(0.0032)	(0.0056)	(0.0074)	(0.0049)	(0.0064)	(0.0081)	(0.0018)	(0.0033)	(0.0037)
Worthless	0.0054	0.0089	-0.0026	0.0021	0.0034	0.0002	0.0001	0.0042	-0.0032
	(0.0027)	(0.0039)	(0.0058)	(0.0047)	(0.0054)	(0.0056)	(0.0013)	(0.0029)	(0.0036)
Restless	0.0064	0.0126	-0.0186	0.0079	0.0045	0.0241	0.0003	0.0094	-0.0157
	(0.0069)	(0.0092)	(0.0180)	(0.0052)	(0.0092)	(0.0179)	(0.0029)	(0.0048)	(0.0069)
Nervous	0.0135	0.0164	-0.0112	0.0043	0.0150	-0.0125	-0.0021	0.0060	-0.0119
	(0.0062)	(0.0099)	(0.0133)	(0.0047)	(0.0075)	(0.0150)	(0.0027)	(0.0039)	(0.0063)
Effort	0.0050	0.0101	-0.0071	-0.0007	0.0056	0.0116	0.0035	0.0084	-0.0156
	(0.0066)	(0.0094)	(0.0156)	(0.0058)	(0.0055)	(0.0121)	(0.0026)	(0.0039)	(0.0062)
Any of the above	0.0178	0.0124	-0.0181	0.0126	0.0142	0.0041	0.0014	0.0140	-0.0157
	(0.0084)	(0.0111)	(0.0208)	(0.0073)	(0.0114)	(0.0222)	(0.0042)	(0.0057)	(0.0080)

Notes: All estimates are from OLS regressions. Models include controls for education, income relative to poverty line, age, race, marital status, employment status, and fixed effects for MSA and year-specific quarter of interview. Samples include males aged 24-59. Standard errors adjusted for non-independence of observations within MSAs are in parentheses.

Table 2C. Estimated effect of MSA unemployment rate on self-reported health, by race.

	African-American	Hispanic	White
Health "worse"	0.0033	0.0020	0.0080
	(0.0049)	(0.0062)	(0.0032)
Health "better"	-0.0124	-0.0055	0.0021
	(0.0080)	(0.0091)	(0.0054)
Sample size	4,077	4,218	16,110

Notes: All estimates are from OLS regressions. MSA unemployment rate is measured annually, rather than quarterly, to match the phrasing of the relevant question. Models include controls for education, income relative to poverty line, age, marital status, employment status, and fixed effects for MSA and year-specific quarter of interview. Samples include males aged 24-59. Standard errors adjusted for non-independence of observations within MSAs are in parentheses.

Table 2D. Estimated effect of MSA unemployment rate on selected health behaviors, by race.

	African American	Hispanic	White
Current smoker	0.0300 (0.0111)	0.0003 (0.0083)	-0.0044 (0.0056)
Any days with 5+ drinks	-0.0054 (0.0117)	0.0030 (0.0099)	0.0003 (0.0077)
Days with 5+ drinks	0.7797 (1.9285)	-0.6479 (0.8988)	-0.7148 (0.6628)
Moderate exercise, any times per week	0.0056 (0.0124)	-0.0182 (0.0163)	0.0091 (0.0084)
Moderate exercise, 3+ times per week	0.0048 (0.0106)	-0.0169 (0.0156)	-0.0012 (0.0078)
Moderate exercise, 5+ times per week	-0.0041 (0.0100)	-0.0023 (0.0139)	-0.0001 (0.0077)
Vigorous exercise, any times per week	-0.0252 (0.0189)	-0.0106 (0.0168)	0.0010 (0.0096)
Vigorous exercise, 3+ times per week	-0.0243 (0.0155)	-0.0076 (0.0105)	0.0019 (0.0080)
Vigorous exercise, 5+ times per week	-0.0213 (0.0117)	0.0073 (0.0095)	-0.0009 (0.0054)
Strength training, any times per week	0.0051 (0.0138)	-0.0032 (0.0106)	0.0035 (0.0063)
Strength training, 3+ times per week	-0.0033 (0.0101)	-0.0070 (0.0065)	-0.0039 (0.0061)
Strength training, 5+ times per week	0.0006 (0.0052)	-0.0044 (0.0066)	0.0032 (0.0049)

Notes: All estimates are from OLS regressions. Models include controls for education, income relative to poverty line, age, marital status, employment status, and fixed effects for MSA and year-specific quarter of interview. Smoking equations also include state-level excise tax on cigarettes as a covariate. Samples include males aged 24-59. Standard errors adjusted for non-independence of observations within MSAs are in parentheses.

Table 3A. Estimated effect of MSA unemployment rate on weight-related health, by education level.

	Less than B.S.	B.S. or higher
Log of BMI	0.0001	0.0050
	(0.0017)	(0.0026)
BMI	0.0120	0.1480
	(0.0555)	(0.0732)
BMI between 18.5 and 25	-0.0003	-0.0126
	(0.0056)	(0.0103)
BMI between 20 and 25	-0.0001	-0.0138
	(0.0054)	(0.0101)
BMI between 18.5 and 30	-0.0008	-0.0106
	(0.0050)	(0.0078)
BMI between 20 and 30	-0.0006	-0.0119
	(0.0054)	(0.0072)
Underweight (BMI≤18.5)	0.0002	-0.0007
	(0.0007)	(0.0013)
Overweight (BMI≥25)	0.0001	0.0133
	(0.0053)	(0.0102)
Obese (BMI≥30)	0.0007	0.0114
	(0.0049)	(0.0075)
Sample size	17,951	8,307

Notes: All estimates are from OLS regressions. Models include controls for education (relevant categories only), income relative to poverty line, age, race, marital status, employment status, and fixed effects for MSA and year-specific quarter of interview. Samples include males aged 24-59. Standard errors adjusted for non-independence of observations within MSAs are in parentheses.

Table 3B. Estimated effect of MSA unemployment rate on psychological well-being, by education level.

	Less than bachelor's degree			Bachelor's degree or higher		
	Most	Some	Never	Most	Some	Never
Sad	0.0038	0.0070	-0.0016	0.0010	0.0069	-0.0104
	(0.0027)	(0.0050)	(0.0075)	(0.0024)	(0.0039)	(0.0083)
Hopeless	0.0052	0.0128	-0.0213	0.0014	0.0049	-0.0073
	(0.0018)	(0.0033)	(0.0035)	(0.0017)	(0.0028)	(0.0049)
Worthless	0.0025	0.0076	-0.0047	0.0016	0.0056	-0.0052
	(0.0019)	(0.0031)	(0.0032)	(0.0020)	(0.0032)	(0.0048)
Restless	0.0032	0.0094	-0.0058	0.0029	0.0067	-0.0072
	(0.0033)	(0.0050)	(0.0089)	(0.0028)	(0.0056)	(0.0099)
Nervous	0.0050	0.0136	-0.0120	-0.0029	-0.0003	-0.0105
	(0.0026)	(0.0040)	(0.0056)	(0.0027)	(0.0056)	(0.0083)
Effort	0.0053	0.0106	-0.0089	0.0073	0.0080	-0.0136
	(0.0029)	(0.0048)	(0.0070)	(0.0034)	(0.0058)	(0.0096)
Any of the above	0.0087	0.0137	-0.0096	0.0039	0.0108	-0.0093
	(0.0038)	(0.0055)	(0.0100)	(0.0042)	(0.0082)	(0.0121)

Notes: All estimates are from OLS regressions. Models include controls for education (relevant categories only), income relative to poverty line, age, race, marital status, employment status, and fixed effects for MSA and year-specific quarter of interview. Samples include males aged 24-59. Standard errors adjusted for non-independence of observations within MSAs are in parentheses.

Table 3C. Estimated effect of MSA unemployment rate on self-reported health, by education level.

	Less than B.S.	B.S. or higher
Health "worse"	0.0038	0.0009
	(0.0039)	(0.0034)
Health "better"	-0.0075	0.0077
	(0.0050)	(0.0069)
Sample size	18,319	8,439

Notes: All estimates are from OLS regressions. MSA unemployment rate is measured annually, rather than quarterly, to match the phrasing of the relevant question. Models include controls for education (relevant categories only), income relative to poverty line, age, race, marital status, employment status, and fixed effects for MSA and year-specific quarter of interview. Samples include males aged 24-59. Standard errors adjusted for non-independence of observations within MSAs are in parentheses.

Table 3D. Estimated effect of MSA unemployment rate on selected health behaviors, by education level.

	Less than a B.S.	B.S. or higher
Current smoker	0.0037 (0.0052)	0.0066 (0.0066)
Any days with 5+ drinks	-0.0044 (0.0068)	0.0114 (0.0096)
Days with 5+ drinks	-0.8088 (0.7527)	0.4465 (0.5112)
Moderate exercise, any times per week	0.0018 (0.0089)	0.0013 (0.0097)
Moderate exercise, 3+ times per week	0.0008 (0.0087)	-0.0106 (0.0085)
Moderate exercise, 5+ times per week	0.0041 (0.0071)	-0.0032 (0.0084)
Vigorous exercise, any times per week	-0.0041 (0.0128)	-0.0043 (0.0088)
Vigorous exercise, 3+ times per week	0.0008 (0.0106)	-0.0104 (0.0093)
Vigorous exercise, 5+ times per week	0.0022 (0.0079)	-0.0092 (0.0084)
Strength training, any times per week	0.0046 (0.0074)	0.0041 (0.0070)
Strength training, 3+ times per week	0.0004 (0.0058)	-0.0103 (0.0074)
Strength training, 5+ times per week	0.0020 (0.0035)	0.0010 (0.0089)

Notes: All estimates are from OLS regressions. Models include controls for education (relevant categories only), income relative to poverty line, age, marital status, employment status, and fixed effects for MSA and year-specific quarter of interview. Smoking equations also include state-level excise tax on cigarettes as a covariate. Samples include males aged 24-59. Standard errors adjusted for non-independence of observations within MSAs are in parentheses.

Table A1. Selected characteristics, by percentile of predicted employment distribution.

	10th percentile	10th-90th percentiles	90th percentile
Employed	0.587	0.907	0.976
	(0.492)	(0.290)	(0.153)
Non-white	0.437	0.243	0.093
	(0.496)	(0.429)	(0.290)
Less than high school	0.496	0.160	0.011
	(0.500)	(0.367)	(0.104)
High school or less	0.703	0.396	0.128
	(0.457)	(0.489)	(0.334)
Bachelor's degree or higher	0.096	0.301	0.607
	(0.295)	(0.459)	(0.488)
Age	40.731	40.504	34.709
	(10.323)	(9.776)	(5.482)
Married	0.326	0.500	0.943
	(0.469)	(0.500)	(0.232)
Living with partner	0.064	0.061	0.032
	(0.244)	(0.239)	(0.176)
Separated or divorced	0.210	0.160	0.015
	(0.407)	(0.367)	(0.120)
Never married	0.359	0.264	0.010
	(0.480)	(0.441)	(0.101)
Sample size	2,715	21,728	2,715

Notes: The column labeled "10th percentile" reflects those in the lowest decile of the predicted employment status distribution, "90th percentile" reflects individuals in the highest decile and "10th-90th percentile" reflects those in the intermediate eight deciles.

References

Acs, Gregory and Douglas Wissoker (1991). "The impact of local labor market markets on the employment patterns of young inner-city males", Unpublished, Urban Institute.

Bartik, Timothy J. (1991). *Who benefits from state and local economic development policies?* Upjohn Institute for Employment Research. Kalamazoo, Michigan.

Bartik, Timothy J. (1993a). "The effects of local labor markets on individual labor market outcomes for different demographic groups and the poor", W. E. Upjohn Institute Working paper 93-23.

Bartik, Timothy J. (1993b). "Economic development and black economic success", Upjohn Institute Technical Report No. 93-001.

Bartik, Timothy J. (1994). "The effects of metropolitan job growth on the size and distribution of family income", *Journal of Regional Science*, 34(4): 483-501.

Bartik, Timothy J. (1996). "The distributional effects of local labor market demand and industrial mix: Evidence from panel data", *Journal of Urban Economics,* 40(2): 150-178.

Blank, Rebecca (1989). "Disaggregating the effect of the business cycle on the distribution of income", *Economica*, 56: 141-163.

Bound, John and Harry Holzer (1993). "Industrial shifts, skill levels and the labor market for white and black males", *The Review of Economics and Statistics*, 75(3): 387-396.

Bound, John and Harry Holzer (1995). "Structural changes, employment outcomes and population adjustments among whites and blacks, 1980-1990", Institute for Research on Poverty Discussion Paper #1057-95.

Brenner, M. Harvey and Anne Mooney (1983). "Unemployment and health in the context of economic change", *Social Science Medicine*, XVII: 1125-1138.

Catalano, Ralph and David Dooley (1983). "Health effects of economic instability: A test of the economic stress hypothesis", *Journal of Health and Social Behavior*, 24: 46-60.

Catalano, Ralph (1991). "The health effects of economic insecurity", *American Journal of Public Health*, 81(9): 1148-1152.

Cawley, John and Kosali I. Simon (2003). "Health insurance coverage and the macroeconomy", NBER Working paper #10092.

Cohen, Sheldon, D.A. Tyrell and A.P. Smith (1991). "Psychological stress and susceptibility to the common cold", *New England Journal of Medicine*, 325: 606-612.

Dooley, David, Ralph Catalano, Karen Rook and Seth Serxner (1989). "Economic stress and suicide: Multilevel analysis", *Suicide and Life Threatening Behavior*, XIX: 321-336.

Dooley, David, Ralph Catalano and Karen Rook (1988). "Personal and aggregate unemployment and psychological problems", *Journal of Social Issues*, 44(4): 107-123.

Fenwick, Rudy and Mark Tausig (1994). "The macroeconomic context of job stress", *Journal of Health and Social Behavior*, 35(3): 266-282.

Freeman, Richard (1991). ""Employment and earnings of disadvantaged young men in a labor shortage economy", In *The Urban Underclass*, eds. C. Jencks and P. Peterson. Washington, DC: Brookings Institution Press.

Hammermesh, Daniel S. and Neal M. Soss (1974). "An economic theory of suicide", *Journal of Political Economy*, 82: 83-98.

Holzer, Harry (1991). "Employment, unemployment and demand shifts in local labor markets", *The Review of Economics and Statistics*, 73(1): 25-32.

Hoynes, Hilary W. (2000). The employment, earnings and income of less skilled workers over the business cycle. In *Finding jobs: Work and welfare reform* (eds: Rebecca Blank and D. Card), New York: Russell Sage Foundation, 23-71.

Kessler, Ronald C. et al. (2003). "Screening for serious mental illness in the general population", *Archives of General Psychiatry*, 60: 184-189.

Philipson, Tomas and Richard A. Posner (1999). "The long-run growth in obesity as a function of technological change", NBER Working Paper #7423.

Ruhm, Christopher J. (1995). "Economic conditions and alcohol problems", *Journal of Health Economics*, 14(5): 583-603.

Ruhm, Christopher J. (2000). "Are recessions good for your health?", *Quarterly Journal of Economics*, 115(2): 617-650.

Ruhm, Christopher J. (2001). "Economic expansions are unhealthy: Evidence from microdata", NBER working paper #8447.

Ruhm, Christopher J. and William E. Black (2002). "Does drinking really decrease in bad times?", *Journal of Health Economics*, 21(4): 659-678.

Ruhm, Christopher J. (2003). "Healthy living in hard times", NBER working paper #9468.

CHAPTER 5

CONCLUSION

The chapters of this dissertation offer empirical evidence on the impact of public policies and economic conditions on the health and human capital formation of individuals. In what follows, I recount my major findings.

Chapter 2 examined an unintended consequence of higher taxation of cigarettes. In particular, I investigated whether recent historically-large cigarette tax increases led to weight gains among smokers. Using repeated cross-sectional data, I found that higher cigarette taxes are associated with an increase in the body mass index of female smokers, but found no similar gain for their male counterparts. Since weight gains among smokers, who weigh less than non-smokers, may not represent poorer health, I also investigated possible distributional impacts. I found increases in clinical obesity among these women, but no effect on the fraction clinically underweight. My findings raise the possibility that the aggregate health benefits from tax-induced smoking reductions may be smaller than anticipated.

Chapter 3 investigated the impact of full-day kindergarten on academic performance. Using longitudinal data, I estimated the impact of full-day kindergarten attendance on standardized test scores in mathematics and reading, as children progressed from kindergarten to first grade. I found that full-day kindergarten has sizeable impacts on student achievement, but the estimated gains were short-lived, particularly for minority children. Given the additional expense of full-day kindergarten, information regarding the size, distribution and duration of gains should be of great interest to policymakers.

Chapter 4 examined the impact of local labor market conditions on health. Using repeated cross sectional data, I estimated the relationship between local labor market conditions and several measures of health and health behaviors for a sample of men living in the fifty-eight largest metropolitan statistical areas (MSAs) in the United States. For those men least likely to be employed, I found consistent evidence of a procyclical relationship for body weight and psychological well-being, but no systematic relationship for a variety of health behaviors including cigarette smoking, heavy alcohol consumption, and various forms of physical exercise. Consistent with these findings, I presented evidence that worsening labor market conditions lead to weight gains and reduced psychological well-being among African American men and lower psychological well-being among less educated males.