ESSAYS IN LABOR AND HEALTH ECONOMICS

ECONOMIC EFFECT OF OBESITY ON WAGES AND ITS IMPACT OVER TIME

by

Nayla Gebran Dahan

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics

Fall 2010

ESSAYS IN LABOR AND HEALTH ECONOMICS

ECONOMICS EFFECTS OF OBESITY ON WAGES AND ITS IMPACT OVER TIME

by

Nayla Gebran Dahan

Approved: ____

Saul D. Hoffman, Ph.D.
Chair of the Department of Economics

Approved: ____

Conrado M. Gempesaw, Ph.D.
Dean of the College of Business and Economics

Approved: ____

Charles G. Riordan, Ph.D.
Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Charles R. Link, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Evangelos M. Falaris, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

David Black, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Simon Condliffe, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

Many thanks to my dissertation committee which provided insightful and timely responses to each draft of the dissertation. Thanks to Dr. Falaris for both teaching me my econometrics and also for his helpful comments throughout the writing of the dissertation. Thanks to Professor Black for his comments and time to finalize the dissertation. Thanks also to Dr. Condliffe for his enthusiastic support as well for his helpful comments on both my master thesis and my dissertation. Larry Hotchkiss not only was a helpful resource in Unix, SAS, and STATA portions of the dissertation but he also provided helpful comments in a very friendly spirit. I was very fortunate to have the impacts of all the above professors on my dissertation.

I am deeply appreciative to the University of Delaware in general and to the Economics department in particular for their support, both from academic and financial standpoint. The faculty has been wonderful and has provided me with the background that will allow me to compete in the dynamic labor market that faces UD graduates. I would like to thank Dr. Hoffman, who was a great teacher and a friendly supportive head of department during the six years. Deborah Sharpley has been both an excellent department secretary and a friend.

I would like to express my deep appreciation and love to my family, my father Gebran to whom I dedicate this book. I love you Baba. To my mom Therese, my brothers Mario, Najib, and Nouhad who have faithfully provided love and affection to me over the years. To my brother Fuad who is the reason for me being here.

Big thanks to my friends Jim Marklam and his wife Barbara for their support and without forgetting that Jim edited all my writing, thank you Jim. To my friends who all of them became part of my family. Their encouragement and support have been sincere and very helpful during my stressful moments.

My deepest appreciation is for my advisor, Dr. Charles Link. His support, help, dedication, and competence are outstanding. This dissertation is dedicated to him too. I am thankful that he served on my committee and that he played such an important role in my career as a graduate student. You have been a very important part of my life for four years. I will miss you.

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

This dissertation presents estimates of the effect of weight on wages in the U.S. Several questions are of interest. Do heavier people earn lower wages? Are the effects of weight on wages evenly distributed over the whole range of wages or are the effects concentrated in the lower, middle or upper part of the wage distribution? Do the effects of weight on wages change over time? This dissertation uses two large data sets, the National Longitudinal Surveys of Youth, NLSY79 and NLSY97, and several regression strategies in an attempt to provide answers to these questions. Differences across gender and race are explored.

The key finding is that weight lowers wages for white females. Negative correlations between weight and wages observed for other gender-ethnic groups appear to be due to unobserved heterogeneity. The results also suggested that the weight penalty, if it exists, increases with wages for almost all sub-groups except Black males. Finally, the negative effect of weight appears to have decreased when we compare the weight penalty between two cohorts, NLSY79 cohort and NLSY97 cohort, aged between 19 and 29. More research is needed so that we can gain insights about the

causes of these penalties. It also provides incentives for policy makers to come up with policies that will help people attain and maintain a healthy weigh.

ECONOMICS EFFECTS OF OBESITY ON WAGES AND ITS IMPACT OVER TIME

Chapter 1

INTRODUCTION

In the past few decades, obesity rates among Americans have skyrocketed. Finkelstein, Cohen, and Dietz (2009) report that, between 1998 and 2006, the prevalence of obesity (body mass index (BMI) greater than 30) increased by 37 percent. Ruhm (2007) examines past patterns and projects future rates of obesity among US adults based on trends in BMI. Ruhm predicts that by the year 2020, 77.6% of men will be overweight and 40.2% will be obese; his corresponding forecasts for women are 71.1% and 43.3%. In 2008, obesity was responsible for 9.1 percent of annual medical expenditures, compared with 6.5 percent in 1998 (Finkelstein, Cohen, and Dietz (2009)). Moreover, in 2008, obesity added more than $27 billion to Medicare and Medicaid costs, which are funded by taxpayer dollars, and contributed $49 billion to private health insurer costs (Cawley (2010)).

Given the accompanying high medical expenditures, stress on the health care system, and the lost productivity due to disability, illness, premature mortality, lower morale and discrimination, the large growth predicted in severe obesity represents a major public health challenge (HR magazine March 2004). Therefore, in order to limit these costs, it makes sense for society to reduce and prevent obesity. Actually, public health officials in the United States have become increasingly alarmed about the growing incidence of obesity, in part because recent research indicates that societal costs of obesity exceed those of cigarette smoking or alcoholism (Sturm 2002). Therefore, more research about obesity is needed to help guide policy intervention.

In this dissertation, my object is to answer questions that have not been answered before. I investigate the relationship between weight and wages across different gender and race groups. In particular I answer the question: does weight affects wages? This dissertation examines the robustness of previous results concerning the effect of weight on wages using the econometric method of censored regressors (not used in the past to detect the effects of weight on wages) and larger data sets, the National Longitudinal Surveys of Youth, NLSY79, (from 1981 to 2006) which have also not been used in the past to detect this effect. Since the rate of obesity in the United States is increasing, it is important to understand not only the level of the wage penalty but also the change in the wage penalty over time (between cohorts) (to my knowledge no previous researcher has used NLSY79 and NLSY97 to detect this change). Differences across wage brackets are also investigated to see if there are wage brackets that deserve more research (these differences too have not been studied before).

In brief, using NLSY79 I find that weight affects wages mostly for white women. This is the most robust result across data sets and econometric methods. I also find some evidence that the weight penalty is more severe for higher paid occupations. Finally, when comparing the effect of weight on wages between the NLSY79 and NLSY97 cohorts, I find that the wage penalty has decreased over time.

My empirical study starts in chapter 4 by replicating Cawley (2004) with the same data, variables, and methods using the NLSY79 data set from 1981 to 2000. In the next step, I use censored regressors to detect the same effect across different gender and race groups and to check the robustness of the results. I finish by using quantile regressions to see if there is any group of people worth further study. Chapter 4 helps answer the questions: does weight affect wages and are the effects of weight on wages the same across different wage brackets?

Chapter 5 is a replication of Chapter 4 but with a larger data set, the NLSY79 data from 1981 to 2006 (three more rounds are included in the data:

(2002, 2004, and 2006). Chapter 5 investigates the robustness of Chapter 4's results but with an older population.

Chapter 6 answers the question: does the wage penalty change over time? To answer this question, I examined two cohorts aged between 16 and 29. The first cohort, taken from NLSY79, was born between 1957 and 1964. The first cohort reflects the effect of weight on wages in the 1980s. The second cohort, taken from NLSY97, was born between 1980 and 1984. The second cohort reflects the effect of weight on wages during the 2000s.

The order of the discussion is as follows. Chapter 2 provides a review of the literature concerning the effects of weight on wages. My data, variables, and methods are presented in Chapter 3. Chapter 4 presents the results from using NLSY79 from 1981-2000. Chapter 5 also investigates the effect of weight on wages and the penalty across different wage brackets but expands the NLSY79 data by including three more years, 2002, 2004, and 2006. Finally, Chapter 6 investigates the change in the wage penalty by comparing the effect of weight on wages for two cohorts, NLSY79 and NLSY97, aged between 16 and 29. Chapter 7 concludes and draws recommendations for policymakers.

Chapter 2

SURVEY OF THE LITERATURE

2.1 Introduction

Historically, body image has been important, especially for women. But not until recent years has thinness been so highly associated with physical attractiveness (Banner (1983)). In the second half of the 1800s, thin women were subject to ridicule. Similar to today's wide marketing for services and products to make one thinner, products then promised to add pounds and produce more rounded, more feminine bodies. Nevertheless, in the 1800s, more weight for men also signifies wealth and well being. In the 2000s, there is the thought that thinness (healthy weight with BMI<25) is a sign of elegance, wealth, and beauty.

Recent studies show that excessive body weight, or obesity (defined as BMI > 30), is a handicap to social advancement, especially for women. The consequences of obesity are numerous, both in terms of increased health problems (diabetes, cardiovascular diseases, and hypertension), and an adverse impact on the quality of life. However, the relationship between obesity and lower wages is complex. Obesity might have economic causes if, for example, low wages induce obesity as low wage individuals consume cheaper but more fattening food. Also obesity might have genetic causes, for example, when all family members are obese due to genetic problem. Also obesity might have nongenetic causes, such as individual choices (Cawley 2004).

This chapter surveys the effects of obesity on wages, employment, income, marital status, and occupational prestige (economic outcomes). It is organized as follows: 2.2 describes the results of previous studies; 2.3 presents the dependent and independent variables used throughout the literature; 2.4 presents conceptual issues and a basic model of weight and wages by Cawley (2004); 2.5 discusses methods employed in the related literature.

2.2 Results of Previous Studies

The economic literature on the effects of weight has grown over the past fifteen years. A primary issue is the extent to which increasing weight affects wages.

Register and Williams (1990), using cross-section data from 1982 NLSY79, found a wage penalty for obesity. Their estimated earnings functions for samples of roughly 8,000 males and females indicate that, among the 18-25 year olds studied, obese females earned more than 12% less than comparable non-obese females. They found no significant effect of obesity on earnings for the males studied. The authors noted the need to repeat the analysis for an older sample, since many 18 to 25 year old individuals are in school and therefore have highly variable wages. In their first model, using BMI as of 1981 as the weight variable, Register and Williams show a statistically significant 15 percent penalty on hourly wages for women with a BMI greater than or equal to 30. Using the 1988 BMI as the weight variable, they find a 10 percent penalty on wages. For men, they find an 8 percent penalty using the 1982 BMI and a 3 percent penalty using the 1988 BMI. The primary impact of discrimination, however, is found in the marriage market, as both the probability of being married and spouse's earnings account for a large portion of the difference in economic status between obese and non-obese women. Furthermore, the authors found that women who became obese between 1981 and 1988 seem to be no worse off than women of recommended weight. But those who were obese at younger ages face significant wage penalties.

Loh (1993) examined the wage effect of height and weight in a sample of full time young workers drawn from the 1982 National Longitudinal Survey of Youth (NLSY) cohort and wage changes between 1982 and 1985. His central result is that physical attributes matter in the labor market for both men and women. He finds that height and weight have statistically significant impact on wage growth in the years from 1982 to 1985; wages for obese men and women grew roughly 5 percent less between the two rounds. Concerns remained regarding the endogeneity of weight (Loh did not correct for endogeniety); weight may very well be correlated with unobserved factors in the error term of the wage equation, causing bias.

In a nationally representative sample (NLSY79) of 10,039 randomly selected young people who were 16 to 24 years old in 1981, Gortmaker et al. (1993) studied the relation between being overweight and subsequent educational attainment, marital status, household income, and self-esteem. He found that women who had been overweight had completed fewer years of school, were less likely to be married, had lower household incomes, and had higher rates of household poverty seven years later than the women who had not been overweight, independent of their base-line socioeconomic status and aptitude-test scores. Men who had been overweight were less likely to be married. Moreover he found that being overweight during adolescence has important social and economic consequences which are greater than for many chronic physical conditions.

Averett and Korenman (1996) use a sample of respondents ages 23 to 31 (5,090 women, 4,951 men) from the 1988 survey of the NLSY79 sample to examine wage rate differentials. Their study is a more direct test of discrimination than Register and Williams. Averett and Korenmen broke down income differentials into components related to the job market (wages) and the marriage market (likelihood of being married and spouse's earnings). They find that the effects of obesity on wages for women are significant but that men apparently suffer only mild economic sanctions, if any. This "obesity" effect is most profound for white women and is primarily concentrated in the marriage market. Specifically, differences in marriage probabilities between overweight and normal weight women account for 50 to 95 percent of their lower economic status. However, sister fixed effects wipe out these obesity effects for all models except one (predicting total family income without controls for marital status, children, and age of youngest child).

Averett and Korenman (1999) find that the effect of weight on women's wages differs by race. White obese women earn 17% less than do white women of normal weight. Self-esteem does not explain differences in obesity effects between white and Black women. Obesity affects work largely through the marriage market for white women. Obesity significantly reduces their likelihood of marriage and significantly reduces their spouse's earnings. The authors studied a sample of roughly 1354 non-Hispanic black and 3097 non-Hispanic, non-black, women aged 25 to 33 years in 1990 from the NLSY79.

Behrman and Rosenzweig (2001) examine the relationship between BMI and wages among females in the Minnesota Twins Registry. The drawback of their data is its relatively small sample size; even the OLS coefficient on BMI is not statistically significant when they control for schooling and work experience (N=1,518). When the authors estimate a regression differencing within 808 monozygotic (MZ) twin pairs, the coefficient on weight is not statistically significant.

Cawley (2004) using NLSY79, reports that an increase in weight of two standard deviations (roughly 65 pounds) is associated with a 7% decrease in wages of white women.

Baum and Ford (2004) also used the NLSY79 to investigate how obesity affects wages by gender. They find that both men and women experience a persistent obesity wage penalty over the first decades of their careers. They suggest that variables such job discrimination, health related factors, and/or obese workers' behavior patterns may be the channels through which obesity affects wages.

Brunello and D'Hombres (2006) find that BMI affects wages negatively in Europe, and that the size of this effect is larger for males than for females. The discovered relationship is much stronger on average in countries of the ''olive belt'' of Europe – Spain, Greece, Italy and Portugal – than in countries of the ''beer belt'' (Austria, Ireland, Denmark, Belgium and Finland), and statistically significant only in the former group of countries. This result is consistent with the higher concern about weight expressed by Southern Europeans in a recent Euro barometer survey (European Commission, 2006). This concern could be due to the negative labor market effects of an increase in body weight.

Lempert (2007), using annual surveys (between the 1981 and 2000 surveys of the NLSY79 to estimate the effect of being overweight on hourly wages, finds a continual increase in the wage penalty for overweight and obese white women followed through two decades.

2.3 Dependent and Independent Variables

The dependent variable in most studies is the natural log of real hourly wages.

The independent variable of interest is the Body Mass Index (BMI) which is defined as weight in kilograms divided by the square of height in meters (or [the weight in pounds divided by height in inches squared] multiplied by 703). The Body Mass Index BMI categories used to define overweight and obesity correspond to thresholds recommended by the US Department of Agriculture; the World Health Organization (WHO) and the US National Institutes of Health.

The most commonly used measures of body weight are: 1) BMI as a continuous variable; 2) weight in pounds; and 3) indicator variables for clinical classifications: underweight (BMI under 18.5), overweight (BMI between 25.0 and 29.9), and obese(BMI 30.0 or higher) where the excluded category is healthy weight(BMI between 18.5 and 24.9). Cawley (2004) used all three measures. Averett and Korenman computed BMI at two ages: an average BMI of the 1981 and 1982, and an average of the 1988, 1989 and 1990.

Weight and height are self-reported weight and height of respondents and therefore include some degree of reporting error which may bias coefficient estimates. Cawley (2004) obtains estimates of true weight and height in the NLSY data utilizing coefficients reported in NHANES III, the third National Health and Nutrition Examination Survey conducted in 1988-94. NHAMES III surveyed a nationally representative sample of 33,994 persons aged 2 months and older of which 31,311 also underwent physical examination that included both self-reported numbers and actual values taken from physical examinations. Separately by race-gender groups, the NHAMES III study regressed actual weight on reported weight and its square. Judging by the high R^2 of .995, reported weight and its square are strong predictors of actual weight. The same process was repeated for height, leading to similar results. In NHANES III, the self-reported height and weight of NLSY white females results in underestimation of BMI by an average of 1.0 percent. Cawley multiplies self-reports of both weight and height in the NLSY by the coefficient reported in NHANES III according to race-gender group.

Nevertheless, simply demonstrating that obese workers earn less than their non-obese co-workers begs the question of exactly how obesity adversely affects their career earnings. Since Americans typically gain weight as they age, it is possible, for example, that the observed obesity wage penalty may simply be masking employer discrimination against older workers. In order to distinguish the effects of weight from those of age and time, Cawley (2004) included age and time as linear variables. Cawley controls for current and recent pregnancy; females who are pregnant at the time of the study are dropped from the sample. To control for the effects of recent pregnancy, he included the number of children women have delivered.

Discrimination against women and members of minority groups may also be affected by the differences in work experience, educational achievement, and environmental factors. The variables included by Cawley to control for differences in human capital are: general intelligence, a test of academic ability and achievement, (derived from ten Armed Services Vocational Aptitude Battery tests or Armed Forces Qualifying tests) AFQT, highest grade completed, mother's highest grade completed, and father's highest grade completed. To control for characteristics of employment, Cawley controls for years of actual work experience, job tenure, the respondent's occupation and whether the occupation is white or blue collar, current school enrollment, the county unemployment rate, and whether the respondent is employed part-time. Other regressors included are age, year, indicator variables for marital status, and region of residence.

Baum and Ford (2004) used a set of explanatory variables that are almost the same as Cawley's but they include twelve industry dummy variables to control for wages differences across industries.

2.4 Conceptual Issues and Methods used

Following previous studies, I employ a model of weight and wages for individual i at time t as follows:

$$(1)\ \ln W_{it} = X_{it}\gamma + \varepsilon_{it}$$

Where $\ln W_{it}$ is the natural log of hourly wage, the subscript i is for the individual, t for time, X_{it} is a vector of explanatory variables including the variable weight, γ is a vector of coefficients and ε_{it} is the error term. Prior studies have found a negative correlation between body weight and wages.

Standard OLS estimates yield biased results if BMI and the error term are correlated, as reviewed in detail by Cawley (2004). There are at least six reasons why this is the case. There is potential reverse causality between wage and weight, this would be the case if weight is higher among those with low income because these people may have a higher intake of cheap food rich in fat and sugar. Low status individuals with low levels of income and education may also exercise less. BMI can be measured with error—as we rely on self-reported measures of weight and height (Cawley and Burkhauser, 2006) which can be source for possible biais.

Cawley (2004) reviewed the endogeneity problem in detail. He proposed a model of weight and wages showed in equation (1). He initially assumes ε is mean-zero, constant variance random variable that is uncorrelated with the explanatory variables in X_{it}.

However, there may be unobserved factors that affect wages and are correlated with the explanatory variables in X. Such correlation would exist if, for example, unobserved variables such as ability and motivation are correlated with both obesity and wages. This would be true if those with low motivation and/or low ability were more likely to be obese and earn lower wages. If variables such as ability and motivation are unobserved, then they will be captured by the error term, and the obesity variable and the error term will be correlated. Therefore the OLS estimates will be biased if we do not control for these variables.

Research in behavioral genetics suggests that roughly half of the variation in BMI is due to non-genetic factors, such as individual choices and environment, and part is due to genetic factors. In addition, obesity may be influenced by wages (low wage people eat less expensive foods that are high in calories), especially for adult females. Cawley (2004).

Thus we have two sources of potential endogeneity in weight the first is nongenetic and the second is genetic. We can decompose the error term in equation (2):

$$(2)\quad \varepsilon_{it} = G_{it} + NG_{it} + V_{it}$$

where V_{it} is the residual.

Therefore, if BMI is affected by wages and personal characteristecs:

$$(3)\quad B_{it} = X_{it}\gamma + W_{it}\alpha + Z_{it}\phi + G_{it} + NG_{it} + \xi_{it},$$

where B_{it} is Body Mass Index, X_{it} is the same vector of explanatory variables as in equation (1) without B_{it}, W_{it} represents wages, Z_{it} is the vector of variables that affect BMI but do not have direct effect on wages, G_{it} is the vector of genetic variables, and finally NG_{it} is the influence of nongenetic factors on BMI. ξ_{it} is the residual.

Equation (3) demonstrates the pitfalls of equation (1). If equation (3) produces a significant coefficient for W_{it}, then OLS estimates do not produce consistent estimates of the true effect of BMI on wages.

2.5 Estimation of the effect of obesity on wages

Most studies have shown that white women are the group for which weight has the most statistically significant effect on wages. But there is variation in this effect due to the way various authors estimate and deal with the problem of endogeneity of the BMI. In most studies the samples used were from the NLSY79. The weight and height reports start in 1981. The length of the sample differs from one author to another. The latest data analyzed is 2000, and the lagged BMI variable used differs from one study to another.

In the previous literature, three strategies have been used to adjust for the possibility that weight is endogenous. The first is to replace weight with a lagged value of weight. This strategy is based on the assumption that lagged weight is uncorrelated with the current wage residual, which assumes no serial correlation in the wage residuals for the two periods. While this strategy will remove any contemporaneous effect of wages on weight, it does not deal with the problem that the genetic and nongenetic components of lagged weight may be correlated with the genetic and nongenetic components of current wages.

Gortmaker et al. (1993), Sargent and Blanchflower (1995), and Averett and Korenman (1996) regress current personal income or wages on measures of body weight from seven years earlier. Cawley (2004) also regresses current hourly wages on measures of body weight from seven years earlier. NLSY recorded the self-reported weight of respondents in 1981, 1982, 1985, 1986, 1988, 1989, 1990, 1992, 1993, 1994, 1996, 1998, and 2000. Each study found that the income or wages of young females was lower if they had been overweight or obese in the past. Each study also found little if any evidence of a difference in wages for males based on weight status seven years earlier.

The second strategy used to deal with the endogeneity of weight is to estimate Equation (1) after taking differences with another individual with highly correlated genes (either a same-sex sibling or twin). Based on equation (3), the differenced regressor of interest is:

$$BMI_{1t} - BMI_{2t} = (X_{1t} - X_{2t})\gamma + (W_{1t} - W_{2t})\alpha + (Z_{1t} - Z_{2t})\phi$$

$$+ (G_{1t} - G_{2t}) + (NG_{1t} - NG_{2t}) + (\xi_{1t} - \xi_{2t})$$

The differencing strategy assumes that all unobservable heterogeneity, genetic and nongenetic, is constant within pairs (that is G1=G2 and NG1=NG2) so that all relevant unobserved heterogeneity is differenced away. This strategy also assumes that wages do not influence weight (α=0), so that the differenced weight variable is uncorrelated with the differenced wage residual.

Averett and Korenman (1996), Averett and Korenman (1996), Behrman and Rosenzweig (2001), and Conley and Glauber (2005) use information on siblings and twins to remove the common household effect of both genetic and non-genetic factors. In taking this difference, they eliminate the variance in weight attributable to shared genes or shared environment. However after differencing they are still left with: a) the variance in weight attributable to nongenetic factors unshared by siblings since NG1≠NG2 with sibling pairs. The coefficient on weight estimated by the siblings-differencing procedure of Averett and Korenman is not statistically significant, which is attributable in part to their small sample of siblings (288 sister pairs and 570 brother pairs).

Baum and Ford (2004) start with the assumption that the error term is not correlated with the vector of explanatory variables (including BMI). They estimate the model using OLS without worrying about the endogeneity of the BMI. To take care of the endogeneity of BMI, they use three approaches. In the first approach, they assume that the wage model contains an individual fixed component, v_i, which represents individual-specific unobserved heterogeneity, the wage equation becomes:

$$\ln W_{it} = \beta X_{it} + v_i + \varepsilon_{it},$$

where W_{it} is the wage and X_{it} is the vector of explanatory variables that includes variables measuring obesity, v_i is an individual-specific factor representing unobserved characteristics and ε_{it} is the error term. If there is correlation between X and v_i and if the individual heterogeneity component v_i is unobserved by the researcher, then estimates of the X coefficients will be biased. The fixed-effects technique for dealing with endogeneity uses cases where respondents provide at least two wage observations by taking the difference between wage observations from the same respondent across time:

$$\ln W_{it} - \ln W_{it-1} = \beta(X_{it} - X_{it-1}) + (v_i - v_{it-1}) + (\varepsilon_{it} - \varepsilon_{it-1})$$

Then v_{it} and v_{it-1} cancel each other (assuming that they are constant over time) and the unobserved variable drops out.

The second strategy used by Baum and Ford to deal with endogeneity is to allow the unobserved heterogeneity to change over time but to assume that such heterogeneity at time t is family-specific. That is, the wage contains a component, v_{ft}, which represents family-specific unobserved heterogeneity at time t.

$$\ln W_{it} = \beta X_{it} + v_{ft} + \varepsilon_{it}$$

As before, if correlation between X and v_{ft} exists and if the family heterogeneity component v_{ft} is unobservable to the researcher, then estimates will be biased. To control for this source of bias, this fixed-effects technique uses cases where the respondent has a sibling and takes the difference between wage observations between siblings from the same family at time t:

$$\ln W_{it} - \ln W_{jt} = \beta(X_{it} - X_{jt}) + (v_{ift} - v_{jft}) + (\varepsilon_{it} - \varepsilon_{jt})$$

The third strategy to address the endogeneity problem controls for both individual-specific and family-specific heterogeneity. In this model the wage contains an individual-specific component, v_i, that is intertemporally fixed, as well as a family fixed component, v_{ft}, which may be variable over time: $v_{ift} = v_{jft}$ but $v_{ift} \neq v_{ift-1}$:

$$(\text{Ln } W_{it} - \ln W_{it-1}) - (\ln W_{jt} - \ln W_{jt-1}) = [\beta(X_{it} - X_{jt}) + (v_{it} - v_{it-1}) + (v_{ift} - v_{ift-1}) + (\varepsilon_{it} - \varepsilon_{it-1})] - [\beta(X_{jt} - X_{jt-1}) + (v_{jt} - v_{jt-1}) + (v_{ift} - v_{ift-1}) + (\varepsilon_{jt} - \varepsilon_{jt-1})]$$

The individual-specific, inter-temporally fixed components, $(v_{jt} - v_{jt-1})$ and $(v_{it} - v_{it-1})$, cancel out, as does the variable family-specific component, $(v_{jft} - v_{jft-1}) + (v_{ift} - v_{ift-1})$.

The last strategy used to deal with the endogeneity of weight is to use variables Z as instruments in instrumental variable estimation under the assumption that $Z_{it} \perp \varepsilon_{it}$. Pagan and Davila (1997), Cawley (2000, 2004), and Cawley et al. (2005) use instrumental variables. Pagan and Davila use indicators of health problems, such as self-esteem and family poverty, as instruments. Pagan and Davila (1997) find, using OLS, that obese females earn less than their more slender counterparts and seek to determine whether their OLS estimate is biased. Using a Hausman specification test, they fail to reject the hypothesis that weight is uncorrelated with the error term of the wage equation. However, their test may be called into question because their instruments (family poverty level, health limitations, and indicator variables about self-esteem) are likely correlated with the error term in the wage equation. Given that their IV estimation likely suffers the same kind of bias as OLS, it is not surprising that they fail to reject the hypothesis that OLS and IV coefficients are equal. However, as argued by Cawley (2004), these instruments may not be valid as they are likely to be correlated with earnings. Using a dataset of twins, Behrman and Rosenzweig (2001) rely on a within-twins estimator and select as an instrument the lagged BMI to simultaneously correct for reverse causality and endogeneity.

Cawley (2000, 2004) uses instrumental variables to deal with the endogeneity problem. He attempts to find a set of instruments that are correlated with BMI, but uncorrelated with the error term in the wage equation. The first instrument used is sibling BMI, which is highly correlated with respondent's BMI and is uncorrelated with the residual in the wage equation. To control for the age and gender of the sibling, the author also includes these variables in the set of instruments. The instruments are good predictors of a woman's weight. Finally, Cawley et al. (2005) study the relationship between obesity and earnings in the US and Germany and use the weight of a child or of a parent as instruments. Results from the simplest statistical model indicate that the relationship between weight and wages varies by race and sex: heavier black men earn more than normal-weight black men, while heavier black women as well as both heavier Hispanic men and women earn less than their normal-weight counter-parts. When individual fixed effects are removed to eliminate the influence of time-invariant unobserved factors among respondents on weight and wages, the negative correlation between weight and wages is eliminated for all but white females, casting doubt on the hypothesis that weight plays a causal role in determining wages for the other groups. This result is further confirmed when time-varying unobserved effects are removed by exploiting the correlation of the respondent's BMI with his or her sibling's BMI.

Chapter 3

DATA, VARIABLES, AND METHODS

3.1 Data: National Longitudinal Survey of Youth 1979 (NLSY79)

The data used in the models estimated in chapters 4 and 5 of this dissertation is from the National Longitudinal Survey of Youth 1979 (NLSY79), which was designed to represent the entire population of American youth in 1979 and consists of a randomly chosen sample of 6,111 U.S. civilian youths, a supplemental sample of 5,295 randomly chosen minority and economically disadvantaged civilian youths, and a sample of 1,280 youths on active duty in the military[1].

NLSY79 data are now available from 1979 (Round 1) to 2006 (Round 22). From 1979 through 1994, the NLSY annually interviewed a cohort of 12,686 respondents who were between the ages of 14 and 21 in 1979 (born between 1957 and 1964). Blacks, Hispanics, and economically disadvantaged persons who were not

---

1. Due to funding constraints, some members of the original sample are no longer being interviewed. After 1984 surveys, interviewing ceased for 1,079 members of the military subsample; retained for continued interviewing were 201 respondents randomly selected from the entire military sample. Beginning with the 1991 survey, 1,643 economically disadvantaged white respondents from the supplemental sample are no longer being interviewed.

black and non-Hispanics were over sampled. The last group is referred to as white throughout this study. Since the 1994 survey, the NLSY is conducted every two years. In each survey, the NLSY collected information on each respondent's employment status, wages, and personal characteristics.

Questions about weight were asked in the 1981, 1982, 1985, 1986, 1988, 1989, 1990, 1992, 1993, 1994, 1996, 1998, 2000, 2002, 2004, and 2006. Height was reported in 1981, 1982, and 1985. Height information was not collected after 1985, presumably because individuals were assumed to have attained adult height (Sample members were 20 to 27 in 1985).

Data from sixteen years (1981, 1982, 1985, 1986, 1988, 1989, 1990, 1992, 1993, 1994, 1996, 1998, 2000, 2002, 2004, and 2006) were pooled to create the sample for this study. Observations that have information about their hourly wage, employment, and occupation were included in the sample. Employed and non-employed persons are included in the sample as long as they have information about hourly wage and occupation[2]. Individuals on active duty in the military were deleted from the sample.

Our sample consists of 6,283 women and 6,403 men who were interviewed

---

2. Of 202,976 observations, we dropped 79,810 or 39.32 percent due to missing hourly wages. Missing information about hourly wages is due to attrition or that the respondent was still at school.

each round and for whom we had the requisite height, weight and hourly wage information.

3.2 Dependent Variable

The dependent variable used in this study is the natural log of the hourly wage. Data on respondent's usual earnings (inclusive of tips, overtime, and bonuses but before deductions) have been collected during every survey year for each employer for whom the respondent worked since the last interview date. The amount of earnings, reported in dollars and cents, is coupled with information on the applicable unit of time, such as per hour. The variable, "Hourly Rate of Pay Current/Most Recent Job", identifies the hourly earnings for the job identified as the CPS job, that is, the job that the respondent held currently or most recently[3].

Following Cawley (2004), outliers in the hourly wage earned by respondent at his or her primary job are recoded. If the hourly wage is less than $1 per hour, it is recoded to $1 and if the hourly wage exceeds $500 an hour, it is recoded to $500[4]. Variables measured in dollar terms such as wages (and other dollar measures) are

---

3. Until 1994, the current or most recent employer was called the CPS employer within the NLSY79. Beginning in 1994 CPS job information is simply labeled as job #1(primary job). Wages are measured by the hourly rate of pay at the current job (CPS job).

4. As a results, 763 person-year observations are bottom-coded and 59 person-year observations are top-coded.

adjusted to reflect 2006 dollars.

3.3 Explanatory Variable

3.3.1 Explanatory Variable of Interest: Body Mass Index (BMI)

The variable of primary interest is Body Mass Index (BMI) is a measure of a person's body weight, scaled according to height, defined as weight in kilograms divided by height in meters squared or as weight in pounds divided by height in inches squared multiplied by 703. The recommended (based on associated mortality risk) BMI range is 20-25. We follow convention in referring to persons below the recommended range as "underweight" persons with BMIs, persons with BMIs between 25 and 29 as "overweight," and persons with BMIs of 30 and over as "obese". We must emphasize that the recommended weight refers to a range associated with low mortality risks, and may not correspond to social norms about what might constitute an overweight appearance[4].

---

4. BMI has been criticized for its inability to distinguish between muscle and fat. Some researchers prefer other anthropometric measures such as waist circumference (Sonmez at al., 2003), waist-hip ratio (Dalton et al., 2003), or waist- height ratio (Cox and Whichelow, 1996). Cawley and Burkhauser (forthcoming) have recently recommended the use of bioelectrical impedance analysis (BIA). None of these are available in the NLSY79. The main reason for the wide-spread use of BMI is its ease of calculation since many data sets contain the necessary information on height and weight

We adopt these ranges because they are conventional, widely used, and are a convenient way to classify persons in our samples.

The NLSY79 recorded the self-reported weight of respondents in the 1981, 1982, 1985, 1986, 1988, 1989, 1990, 1992, 1993, 1994, 1996, 1998, 2000, 2002, 2004, and 2006. The weight is recorded in pounds. Height was reported in 1981, 1982, and 1985. Height information was not collected after 1985. Given that respondents were between the ages of 20 and 27 in 1985, height in 1985 was assumed to be respondent's adult height following Cawley (2004). Height was recorded in inches. However, height is collected each year in the data set NLSY97.

These self-reported of weights and heights include some degree of reporting error, which may bias coefficient estimates. The direction of reporting bias is negatively correlated with actual weight: underweight people tend to over report their weight, and overweight people tend to underreport their weight (Cawley and Burkhauser (2006)). In order to correct for this reporting error, true height and weight in the NLSY are predicted using the coefficients reported by Cawley and Burkhauser (2006) which express the relationship between true and reported values using the Third National Health and Nutrition Examination Survey (NHANES III) [5].

---

5. Third National Health and Nutrition Examination Survey (NHANES III) conducted in 1988-94, surveyed a nationally representative sample of 33,994 persons aged 2 months and older, of which 31,311 of those respondents also underwent physical examinations.

Cawley accounts for this under and over reporting as follows. Separately by race and gender groups, actual weight was regressed on reported weight, its square, and on age and its square (Table 6 Regression of Measured Weight in Pounds on Self-Reported Weight in NHANES III, by Race and Gender). Judging by the high R-squared of .995, reported weight and its squared are strong predictors of actual weight. The same process was repeated for height, leading to the similarly significant results (Table 7 Regression of Measured Height in Inches on Self-Reported Height in NHANES III, by Race and Gender). In NHAMES III, self reported height and weight on NLSY-aged white result in underestimated BMI by an average of 1.58 percent, while male calculated BMI is underestimated by an average of 1.0 percent. Cawley assumes that the relationship between true and reported values is the same across datasets. Finally, Self-reported height and weight in the NLSY are then multiplied by the coefficient on the reported values associated with the correct race-gender group in the NHANES III. The fitted values of weight and height, correcting for reporting error, are used throughout my dissertation [6].

We will use three 3 measures of body weight in this dissertation:

1- Body mass index (BMI) as a continuous variable.

2- Weight in pounds (controlling for height in inches)

6. Cawley and Burkhauser provide those coefficients to researchers so they can estimate a more accurate value of height and weight.

3- Indicator variables for the clinical classifications underweight, overweight, and obese, where the excluded category is healthy weight. (The U.S National Institutes of Health and the World Health Organization (WHO) classify BMI as follows: below 18.5 is underweight, between 18.5 and less than 25 is healthy, greater than 25 and less than 30 is overweight and over 30 is obese.

In the NLSY79, weight and or height were missing for 81,013 (28,099 height, 52,914 weight) observations. Those observations were dropped from the sample.

Following Cawley and Burkhauser (2006), I dropped implausible observations with self-reported height under four feet (one observation) or over seven feet (0 observation) or with self-reported weight under 80 pounds (48 observations) or over 900 pounds (11observations).

Table 1 presents a tabulation of the distribution by BMI categories, gender, and race of the sixteen years pooled NLSY79 data set between1981 and 2006. In our sample, about half of the women are in the recommended (19 to 24) BMI range, more than 24 percent are between 25 and 29, almost 22 percent are over 30 or obese, and almost 4 percent were below 19. Results for men are presented in the bottom half of table 1. More than half of men are overweight or obese and less than half are in the recommended weight (normal weight) category.

3.3.2 Other Control Variables

3.3.2.1 Gender

Variables available within the main NLSY79 data set provide information on the sex of each respondent. During interviews, gender was determined by observation and asked directly of respondents only if it was “not obvious” to the interviewer. These observations are subject to a small degree of error from erroneous interviewer observation or recoding and data entry error. Therefore, when using this series of variables, a small number of respondents may appear to “change” sex across surveys Interviews started in 1979 with 6,403 males and 6,283 females for a total of 12,686[7].

3.3.2.2 Race and Ethnicity

The variable ‘Racial/Ethnic’ designates the respondent as “Hispanic,” “Black,” and “nonblack/non-Hispanic” and *provides the basis for weighting* NLSY79 data. As we noted, using variables of race and ethnicity three groups were created: Black, Hispanic, and nonblack/non-Hispanic. The last group is referred to as white throughout this paper, although it is a heterogeneous group. (Table (2) presents the ethnicity by race and ethnicity as presented by the NLSY79 geo-code book)

7. On March 1, 1986, ‘Sex of R’ was changed for 42 cases as a result of inconsistencies generated from interviewer checks. Three additional cases were changed shortly thereafter. The changes included from male to female and from female to male.

(1) “Hispanics” were those who self-identified as Hispanic

2) “Blacks” included those for whom race was coded “Black” and ethnic origin was “non-Hispanic” or those whose ethnic origin was coded Black, Negro, or Afro-American

3) “Nonblack/non-Hispanics” included those whose race was coded “white” or “other” and who did not identify themselves as either Black or Hispanic in answer to the ethnicity question.

We end up with 6 groups: white female, Black female, Hispanic female, white male, Black male, and Hispanic male.

3.3.2.3 Age and Time

Weight tends to rise with age. In order to control the age effect on weight, linear measures of age and time are included as regressors in wage estimation.

The NLSY79 main data files contain a yearly created variable, 'Age of R at Interview Date.’ These created variables are constructed using the 1981 date of birth information coupled with the 1979 birth date for the 491 respondents not interviewed in 1981.

3.3.2.4 Fertility variables

The first three rounds of the NLSY79 (including 1981) have very short fertility sections. In 1982 the fertility data collection was greatly expanded due to additional funding provided by the National Institute of Child Health and Human Development. During this survey, full retrospective information about the respondent's fertility history was collected. In 1987 the fertility section began a new pattern. In odd years, such as 1987, 1989, 1991, and so forth, only a sub-section of the fertility questions were asked.

Weight may be affected by current or recent pregnancy. For this reason, females who are pregnant at the time that they report their body weight are dropped from the sample. Two questions were used to eliminate from the sample women who are pregnant at the time that they report their weight in the NLSY79. First, women were asked whether they were pregnant at the time of the interview. Second, in some years they were also asked whether they had been pregnant at the time of the last interview. To control for recent pregnancy, the set of regressors includes the age of a woman's youngest child and the total number of children to whom she has given birth.

When using censored regressors data, I create 5 dummy variables (have no kids, have kids between 0 and 5, have kids between 6 and 10, have kids 11-18, have kids >18 where have no kids is the excluded variable) to present the variable" number of children to whom she has given birth" to avoid deleting observations with no kids.

3.3.2.5 Human Capital

General Intelligence: is a measure of cognitive ability derived from the Armed Services Vocational Aptitude Battery (ASVAB) which is a special survey administered in 1980 to the 1979 sample of NLSY79 respondents. NLSY79 respondents were selected since they comprised a nationally representative sample of young people born during the period 1957 through 1964. This testing, which came to be referred to as the "Profile of American Youth," was conducted by NORC representatives according to standard *ASVAB* procedural guidelines. Respondents were paid $50 for their participation. Groups of five to ten persons were tested at more than 400 test sites, including hotels, community centers, and libraries throughout the United States and abroad. A total of 11,914 civilian and military NLSY79 respondents (or 94 percent of the 1979 sample) completed this test: 5,766 or 94.4 percent of the cross-sectional sample, 4,990 or 94.2 percent of the supplemental sample, and 1,158 or 90.5 percent of the military sample.

The ASVAB consists of a battery of 10 tests that measure knowledge and skill in the following areas: 1- general sciences, 2- arithmetic reasoning, 3-word knowledge, 4- paragraph comprehension, 5-numerical operations, 6-coding speed, 7-auto and shop information, 8- mathematics knowledge, 9-mechanical comprehension, and 10- electronics information.

Highest grade completed: Information on the highest grade completed has been collected in each survey year.

Mother's highest grade completed: Highest grade completed is available for household family members. Information the highest grade completed by respondent's mother was collected in 1979.

Father's highest grade completed: Information on the highest grade completed by respondent's father was collected in 1979.

3.3.2.6 Characteristics of Employment

Work Experience: The work experience information collected in the NLSY79 provides a relatively complete picture of the respondent's labor force activities between the previous and current interview dates.

A cumulative work experience (defined as weeks of reported actual work experience divided by 50) for each year are calculated by adding the previous work experience for each year to have the variable cumulative work experience in weeks for the year.

Job tenure: The variable series 'Total Tenure" (in weeks) with Employer as of Interview Date Job #1' is reported by the NLSY79 for all years. A cumulative tenure in weeks was reported with an employer from the time the employer was first reported up to the most recent week worked.

Occupation**:** NLSY79 includes information on the occupation of respondents' current/last job. The 3-digit 1970 Census classifications (U.S. Census Bureau 1971) are used to code all job and training questions in the 1979-2000 surveys (Table 2 OCCUPATION (1970 CENSUS 3 DIGIT)). In that span of 30 years, new occupations were created and others became redundant. It was decided to update to the 2000 codes beginning in 2002. For the 2002 survey, the 3-digit 2000 Census codes (U.S. Census Bureau 2000) were used to classify occupations and industries of all jobs reported by respondents. The Census Bureau issued another revision in 2003**,** and these codes were used for the 2006 survey (Table 2 and 3(CENSUS 3 DIGIT, 2000 CODES)).

Following Cawley (2004) all occupations are classified as either white collar or blue collar. I used the same classification as Cawley. White collar workers are those working in sectors described by the U.S. Census as professional, technical, or kindred workers, nonfarm managers and administrators, sales workers, and clerical. A blue collar occupation is a work in sectors described by the U.S. Census as farming, forestry, fishing, craft, repair, operators, fabricators, and laborers.

Current School Enrollment: Data have been collected during each round of the NLSY79 on respondents' current school enrollment status.

County Unemployment Rate: The NLSY79 unemployment rate variables are constructed using state and area labor force data from the May publication of *Employment and Earnings* for the month of March of each survey year. *Employment* *and Earnings* is published by the U.S. Department of Labor, Bureau of Labor Statistics and lists the civilian labor force and number of unemployed persons for every state and selected metropolitan area. The respondent's metropolitan statistical area is assigned based on the county, state, and zip code of current residence. Respondents who are in the military, who are living outside of the United States, or who have invalid geographic data for a given survey year are valid skips on these variables.

The unemployment rate variable is available only in the restricted-use geocode data files.

Part time vs Full time Job: Dummy variables, whether the respondent's job is part time or full time job, were created using the number of hours worked per week by the respondent in his primary job. If weekly hours worked by respondent are less than 20, his job is considered to be a part time job and if weekly hours worked are greater than 20, the respondent's job is considered to be full time job.

3.3.2.6 Other variables

Marital status The marital status of each respondent, whether he or she was married, never married, or other (widowed, divorced, or separated) was collected during the 1981-2006 surveys and is available as a single variable, 'Marital Status.' Dummy variables were created for marital status; never married, married with spouse present, and married with spouse not present, using the latter as the omitted variable.

Region of residence: Four dummy variables (South, West, North Central, and Northeast) (See were created to present the four regions of residence using South as the omitted variable.

Missing Variables: Indicator variables for missing data associated with each regressor, except the weight variables, are also included.

3.4 Sampling Weight

Sampling weights are useful and often essential for obtaining unbiased estimates of univariate population characteristics from sample data. The distribution of variables in an unweighted sample can differ from that on the population from which it was drawn for two reasons. First, individuals may be sampled with unequal probabilities. Second a sample may also differ from the population from which it was drawn because of random chance.

NLS data come from complex longitudinal surveys featuring multiple representative samples. The NLSY oversampled Blacks, Hispanics and poor non-Hispanic whites. The NLS staff creates sets of cross-sectional weights for each cohort and each of its survey rounds. These weights produce group estimations that are demographically representative of each cohort's base-year population when used in tabulations. We weighted survey data with sample weights provided by the NLSY. Using these weights allows users to correct the raw data for the complex survey design in a particular year and would reflect a sample representative of the United States population.

The data used in this research were weighted using the custom weight program designed by the NLSY for multi-year analysis.

3.5 National Longitudinal Survey of Youths 1997 (NLSY97)

The dataset utilized in chapter 6 of this dissertation is taken from National Longitudinal Survey of Youth 1997 (NLSY 97). The survey sample is designed to represent U.S. residents in 1997 that were born during the years 1980 through 1984. The majority of the oldest cohort members (age 16 as of December 31, 1996) were still in secondary school during the first survey round and the youngest respondents (age 12) had not yet entered the labor market. Interviews with these youths have been conducted annually, starting in 1997. Data is available through 2007 (11 rounds).

The original sample includes 8,984 respondents, comprised of 6,748 respondents reflecting the overall racial/ethnic makeup of the US population in 1997, with an over-sampling of 2,236 Black and Hispanic respondents. Similar to the NLSY79, the NLSY97 offers a rich set of variables for each respondent related to employment, labor market outcomes, demographic, family, fertility, and personal characteristics.

Throughout the analysis, NLSY97 data are weighted using the sampling weights from Round 1.

3.6 Conceptual Issues and Econometric Methods

3.6.1 Conceptual Issues

A naïve model that relates wages to weight assumes that wages (W) and weight represented by Body Mass Index, (or BMI, or some other weight related variable) have the following relationship for individual i at time t:

$$(1)\ LnW_{it} = X_{it}\gamma + \varepsilon_{it}$$

LnW_{it} represents the natural log of wages, X_{it} is a vector of explanatory variables that includes BMI or another (other) variable(s) measuring weight that affect wages and ε_{it} is the residual for observation i in time t. If we initially assume ε_{it} has mean-zero, and constant variance, and is a random variable uncorrelated with the explanatory variables in X_{it}, an Ordinary Least Square (OLS) estimate of γ is consistent estimate of the true effect of weight on wages. However, there may be unobserved factors that affect wages and that are correlated with the explanatory variables in X_{it}. Such correlation might exist if, for example, unobserved variables such as ability and motivation are correlated with both weight and wages. This would be true if those with low motivation and/or low ability were more likely to be obese and earn lower wages. If variables such as ability and motivation are unobserved, then they will be captured by the error term, and the obesity variable and the error term will be correlated. Therefore the OLS estimates will be biased if we do not control for these variables.

Research in behavioral genetics suggests that roughly half of the variation in BMI is due to nongenetic factors such as individual choices and environment, while part is due to genetic factors Cawley (2004). As Cawley suggests there are two sources of potential endogeneity in weight. The first is nongenetic and the second is genetic. We can decompose the error term in equation (1) as:

$$(2)\ \varepsilon_{it} = G_{it} + NG_{it} + V_{it}$$

where V_{it} is the residual.

Weight, represented by BMI in the following equation, may in turn be affected by wages and personal characteristics.

$$(3)\qquad B_{it} = X_{it}\gamma + W_{it}\alpha + Z_{it}\phi + G_{it} + NG_{it} + \xi_{it}$$

In Equation (3), B_{it} is BMI, X_{it} is the same vector of explanatory variables without B_{it}, W_{it} represents wages, Z_{it} is the vector of variables that affect BMI but do not have a direct effect on wages, G_{it} is a vector of genetic variables, and finally NG_{it} is the influence of nongenetic factors on BMI. ξ_{it} is the residual of BMI (Cawley

2004). Equation (3) reveals the pitfall of equation (1) and shows that the estimates of the true effect of BMI on log wages through naïve regressions might not be consistent.

Prior studies have found a negative correlation between body weight and wages (Cawley (2004), Gortmaker et al. (1993), Sargent and Blanchflower (1994) and Averett and Korenman (1996)). But this result can be explained in various ways. One possibility is that obesity lowers wages via lower productivity or discrimination (Cawley 2004). It also might be that causality runs in the opposite direction; low wages might cause obesity. This could be true if low income people eat less expensive foods that are high in calories (Cawley 2004). Obese employees may earn less if clients in customer care industries dislike working with obese workers (Baum & Ford 2004). Obese individuals also tend to have higher health care costs. Employers offering health insurance might offer lower wages to the obese with that in mind (Baum & Ford 2004). Finally it is also possible that unobserved variables cause both obesity and low wages (Cawley 2004).

Given the above, when estimating the effect of weight on wages, a major conceptual issue is the potential endogeneity of weight.

3.6.2 Econometric Methods

3.6.2.1 Ordinary Least Square (OLS) and Fixed Effect Models

I investigate the relationship between weight and wages across race-ethnic and sex categories (white female, Black female, Hispanic female, white male, Black male, and Hispanic male) using the same model specifications as Cawley (2004):

$$(1)\ \text{Ln } W_{it} = X_{it}\gamma + \varepsilon_{it}$$

In Equation 1, X_{it} is a vector of explanatory variables for observation i in time t that includes a variable measuring obesity and ε_{it} is the residual for observation i in time t. We initially assume ε_{it} has mean of zero and constant variance and is uncorrelated with the explanatory variables in X. Then an Ordinary Least Square (OLS) estimate of γ can be interpreted as a consistent estimate of the true effect of weight on log wages.

Given the importance of the endogeneity problem, the OLS estimates will be biased if we do not control for these variables mentioned above. Taking advantage of the longitudinal nature of the NLSY79 data, and following Cawley (2004), first I will use BMI lagged of 7 years. Based on previous studies (Gortmaker et al. (1993), Sargent and Blanchflower (1994) and Averett and Korenman (1996)), the endogeneity problem should be less important if one uses an early value of BMI rather than a contemporaneous value. Second a fixed effect model is used to eliminate the other factors that might affect wages and are unobservable such as ability.

First Strategy: OLS with Lagged Weight

My first strategy is to replace the weight variable with its lagged value in the OLS regression. This strategy is based on the assumption that lagged weight is uncorrelated with the current wage residual: weight $_{it-\tau} \perp \varepsilon_{it}$. This assumes no serial correlation in the wage residuals between periods: $\varepsilon_{it-\tau} \perp \varepsilon_{it}$. While this strategy will remove any contemporaneous effect of wages on weight, it does not deal with the problem that the genetic and nongenetic components of lagged weight ($G_{it-\tau}$, and $NG_{it-\tau}$) may be correlated with the genetic and nongenetic components of current wages (G_{it}, and NG_{it}). So although the independence of the lagged weight variables on the contemporaneous wage residual is required, it might not be likely. The contemporaneous wage error term is likely to capture some omitted variable related to both past weight and the contemporaneous wages.

Second strategy: Fixed Effects model

This section, following Greene (2003), describes the basic statistical framework and estimation procedures used in the estimation. A good feature of the NLSY79 data is that individuals are followed from 1979 to 2006. How to exploit this information raises important questions. The first question is, whether 'individual effects' are an important factor in the data underlying our models.

The panel analysis builds on the regression model of the form:

$$(4)\ \text{Ln } W_{it} = \alpha_i + X_{it}\gamma + \varepsilon_{it}$$

where $E[\varepsilon_{it}]=0$ and $Var[\varepsilon_{it}]=\sigma_{it}^2$

LnW_{it} measures the natural log of wage, X_{it} is a vector of explanatory variables that includes variables measuring obesity (not including the constant term $\alpha_{i)}$ and ε_{it} is the residual for observation i in time t. Introduction of α_i allows for potential heterogeneity (individual effects) across individuals. These effects are assumed to be constant over time, t, and represented by α_i. We initially assume ε_{it} is mean-zero, constant variance random variable that is uncorrelated with the explanatory variables in X_{it}. Then an Ordinary Least Square (OLS) estimate of β can be interpreted as a consistent estimate of the true effect of BMI on log wages.

One may find it helpful to think of these individual specific effects as those factors which affect the wage and are unobserved. These variables might include such things as ability and motivation. This would be true if those with low motivation and/or low ability were more likely to be obese and earn lower wages. If variables such as motivation are unobserved, then they will be captured by the error term, and the obesity variable and the error term will be correlated. Therefore the OLS estimates will be biased if we do not control for these variables. This is the well known ‘omitted variable bias’ in econometrics. To solve the problem of coefficients bias coming from ‘individuals effects’, we follow Baum and Ford (2004), and Cawley (2004) who rely on fixed effect estimators to control for unobservable individual effects.

The fixed effects model assumes that there is a separate constant term for each unit (i.e. differences across individuals are captured by differences in the constant term where there is a constant term α_i for each individual):

$$\textbf{(4)}\ \text{Ln } W_{it} = \alpha_i + X_{it}\gamma + \varepsilon_{it}$$

α_i in the fixed effects specification in Eq. (4) is non stochastic and constant over time for each individual. This model is commonly known as the least squares dummy variables (LSDV) model.

3.6.2.2 Censored regressors

Rigobon and Stoker (2007) study issues that arise for estimation of a linear model when a regressor is censored. They discuss the efficiency losses from dropping censored observations, and illustrate the losses for bound censoring. They show that the common practice of introducing a dummy variable to correct for censoring does not correct bias or improve estimation. They show how censored observations generally have zero semi-parametric information, and they discuss implications for estimation. Rigobon and Stoker (2009) show that there are no informational gains from using the censored data. Therefore, the common practice of introducing a dummy variable to deal with the censoring aspect of the regressors either will not help in the estimation, or it will continue to bias the coefficients.

Following Rigobon and Stoker (2009), I investigate the relationship between weight and wages across race-ethnic and sex categories using censored regressors (deleted all the missing values of the regressors) and compare the results with noncensored regressors (including the dummy variable for missing values). This will allow us to determine whether the Cawley results are significantly altered if we use the Rigobon and Stoker approach.

Weight variables from the NLSY79 sample are used in all estimation described in this paper. T-statistics reported for OLS and IV for both censored and noncensored regressors reflect robust standard errors that are calculated with clustering to account for correlations in the error terms for each individual over time.

3.6.2.3 Quantile Regression

We say that a student scores at the τth quantile of a standardized exam if he performs better than the proportion τ of the reference group of students and worse than the proportion (1-τ). Thus, half of students perform better than the median student and half perform worse. Similarly the quintiles divide the population into five parts, the deciles into ten parts. The quantiles refer to the general case. Quantile regression as introduced by Koenker and Bassett (1978) seeks to extend these ideas to the estimation of conditional quantile function-models in which quantiles of the conditional distribution of the response variable are expressed as functions of observed covariates.

Ordinary least-squares regression models the relationship between one or more covariates X and the conditional mean of a response variable Y given X = x. With OLS method, we abandon the idea of estimating separate means for grouped subsets of data, and we estimate the parameters of the model on the assumption that they are linear. In contrast, quantile regression models the relationship between X and the conditional quantiles of Y given X = x, so it is especially useful in applications where the relationship between Yand X might appear only with extreme values. Quantile regression also provides a more complete picture of the conditional distribution of Y given X = x when both lower and upper or all quantiles are of interest, as in the analysis of wages where both lower (low wages) and upper (high wages) quantiles are closely watched.

Are the effects of weight on wages the same throughout the distribution of wages? For instance, are the effects concentrated in the lower part of the wage distribution? There may be sub-groups worth analyzing further. Quantile regressions are used to investigate the effect of weight on wages across race-ethnic and sex categories dividing the population in five parts: 20^{th} percentile, 40^{th} percentile, 50^{th} percentile, 60^{th} percentile, and 80^{th} percentile.

3.7 Conclusion

The first empirical work (chapter 4) presents all the regressions described above using NLSY79 (1981-2000) data. Chapter 5 expands the study and uses a larger data set NLSY79 with three more rounds (1981-2000, plus 2002, 2004, and 2006) to produce more precise estimates. The innovation in this dissertation is to use censored regressors, quantile regressions and finally a larger data set NLSY79 (1981-2006) that includes three more rounds. Nevertheless, Chapter 6 compares if the effect of weight on wages changed over time. This will be done using data set NLSY97 to compare the fluctuation of the penalty over time. Nobody to my knowledge has used the data set NLSY97 to detect the effect of weight on wages.

Table 3.1

Percentage Distribution of Body Mass Index, Sample 1981-2006

Females

Women	Full Sample	White	Black	Hispanic
Under weight or BMI< 19	3.88%	4.80%	2.66%	2.95%
Normal weight or BMI 19-24	49.65%	58.39%	37.31%	42.22%
Overweight or BMI 25-29	24.60%	21.46%	27.43%	29.82%
Obese or BMI 30+	21.87%	15.35%	32.60%	25.00%

Males

Men	Full Sample	White	Black	Hispanic
Under weight or BMI< 19	1.36%	1.28%	1.57%	1.24%
Normal weight or BMI 19-24	44.83%	46.77%	46.66%	36.12%
Overweight or BMI 25-29	35.58%	35.97%	32.71%	38.87%
Obese or BMI 30+	18.24%	15.98%	19.06%	23.77%

Table 3.2

OCCUPATION (1970 CENSUS 3 DIGIT)

WHAT KIND OF WORK WERE YOU DOING FOR THIS JOB? WHAT WERE YOUR MOST IMPORTANT ACTIVITIES OR DUTIES?

1 TO 195: 001-195 Professional, Technical and Kindred

201 TO 245: 201-245 Managers, Officials and Proprietors

260 TO 285: 260-285 Sales Workers

301 TO 395: 301-395 Clerical and Kindred

401 TO 575: 401-575 Craftsmen, Foremen and Kindred

580 TO 590: 580-590 Armed Forces

601 TO 715: 601-715 Operatives and Kindred

740 TO 785: 740-785 Laborers, except Farm

801 TO 802: 801-802 Farmers and Farm Managers

821 TO 824: 821-824 Farm Laborers and Foreman

901 TO 965: 901-965 Service Workers, except Private Household

980 TO 984: 980-984 Private Household

0: 00 None

990: 990 Same as Present Job

995: 995 Did not work

996: 996 Never Worked

Table 3.3

OCCUPATION (CENSUS 3 DIGIT, 2000 CODES) (ALL) 2000 CENSUS CODE FOR OCCUPATION – EMPLOYER

2000 CENSUS CODE FOR OCCUPATION

1 TO 43: Management

50 TO 95: Business and Financial Operations

100 TO 124: Computer and Mathematical

130 TO 156: Architecture and Engineering

160 TO 196: Life, Physical, and Social Services

200 TO 206: Community and Social Services

210 TO 215: Legal

220 TO 255: Education, Training, and Library

260 TO 296: Arts, Design, Entertainment, Sports, and Media

300 TO 354: Healthcare Practitioners and Technical

360 TO 365: Healthcare Support

370 TO 395: Protective Service

400 TO 416: Food Preparation and Serving Related

420 TO 425: Building and Grounds Cleaning and Maintenance

430 TO 465: Personal Care and Service

470 TO 496: Sales and Related

500 TO 593: Office and Administrative Support

600 TO 613: Farming, Forestry, and Fishing

620 TO 694: Construction and Extraction

700 TO 762: Installation, Repair, and Maintenance

770 TO 896: Production

900 TO 975: Transportation and Material Moving

980 TO 983: Militar

Chapter 4

EFFECTS OF OBESITY ON WAGES: 1981-2000

The objective of this chapter is to estimate the effects of obesity on wages of persons in the U.S. based on the NLS79 data. The outline of the chapter is as follows. Section 4.1 is an introduction to the issues. Section 4.2 discusses the important empirical results (by gender and ethnicity) regarding weight and its impact on wages. In this section we replicate the work of Cawley (2004); and then test the sensitivity of his results for the way he handles missing data in the regressorsin section. In the last part of this section we estimate quantile regressions to see if the wage penalties associated with obesity are concentrated in particular parts of the wage distribution. Section 4.3 is a summary of the chapter. No one to my knowledge has used the method of censored regressors method, suggested by Rigobon and Stoker (2009), to estimate the effect of weight on wages. Another contribution to the literature is that we also investigate whether weight affects all wage brackets the same way by using quantile regressions.

4.1 Introduction

We examine the impact of weight on wages in the U.S using National Longitudinal Survey of Youth (NLSY79), a panel survey, begun in 1979, designed to represent young males and females in the U.S. Any such study must confront the fact that a correlation between BMI and wages needs a causal relationship running from the BMI to wages. The uncovered correlation could in fact reflects both that body weight affects wages and/or that wages affect body weight. We follow Cawley (2000, 2004) and Cawley et al. (2005), and replace the contemporaneous weight with its 7-years lagged value in the OLS regression and second by relying on fixed effect estimators to control for unobserved individual effects.

Moreover this chapter uses an alternative method of handling missing data in the regressors, censored regressors (Rigobon and Stoker (2009)) to examine the robustness of previous results of the effect of weight on wages.

This research finds that excessive weight lowers wages for white females. OLS estimates for white women indicate that a difference in weight of two standard deviations (roughly 65 pounds) is associated with a difference in wage of 9 percent. In absolute value, this is equivalent to the effect of roughly one and half years of education or three years of work experience. Negative correlations between weight and wages observed for other gender-ethnic groups appear to be due to unobserved heterogeneity. Also we found some evidence that the wage penalty for obesity for some gender-ethnic groups is higher at higher wages.

The starting point of my empirical study is replicating Cawley (2004) with the same data, variables, and methods. In my next step, I use the econometric model, censored regressors, to detect the same effect across different gender and race groups. I conclude by using quantile regressions to see if there is a group of people worth further study.

The outline of this chapter is as follows. Section 4.2 presents the empirical results where section 4.2.1 will discuss the descriptive statistics for the two samples underlying our results. Section 4.2.2 presents Cawley replication. Section 4.2.3 discusses the results of the analysis using data with censored regressors and 4.2.3 presents the quantile regressions with censored regressors. I conclude in section 4.3.

4.2 Empirical Results

4.2.1 Descriptive Statistics

Appendix Tables 4A1, 4A2, 4A3, and 4A4 provide summary statistics by gender. Table 4A1 and 4A2 provide descriptive statistics of the samples of females and males respectively, replicating Cawley 2004. That is, the sample includes persons who have missing data in the regressors. Tables 4A3 and 4A4 present the descriptive statistics for the samples of females and males respectively after observations are deleted that have any missing data for a right hand side variable. Let us compare the means for some key variables by gender. The mean age for all the samples is age 29.

First let us examine means for females. For the most part the means are fairly close for both samples. In both tables the mean BMI is around 25, height is 64 inches, and weight is just less than 150 pounds. The exceptions are for wages, mother's years of education, father's years of education, and test of general intelligence. The mean wage (2000 dollars) in the censored sample is $11.96 while in the sample including observations with missing data the mean is $11.56. Means for mother's highest grade completed, father's highest grade completed and general intelligence were for the Cawley sample (censored regressors sample) respectively 10.43 (11.12), 9.63 (11.09) and 38.8 (43.63).

As with the females, the means for males are fairly close in both samples. In both tables the mean BMI is around 25.9, height is 69.7 inches, and weight is about 180 pounds. The exceptions are for wages, mother's years of education, father's years of education, test of general intelligence, and age of the youngest child. The mean wage (2000 dollars) in the censored sample is $14.97 while in the sample including observations with missing data the mean is $14.43. Means for mother's highest grade completed, father's highest grade completed, and general intelligence were for the Cawley sample (censored regressors sample) respectively 10.17 (11.11), 9.52 (11.09) and 38.8 (43.63).

4.2.2 Replication of Cawley (2004)

Tables 4.1 and 4.2 present a replication of Cawley (2004). Indicator variables for missing data associated with each regressor, except the weight variables, are included. Only the coefficients for various measures of weight are included in the tables although various socioeconomic and demographic variables are included (variables included in the regressions are indicated at the bottom of each tables and are defined in section 3.3.1 and 3.3.2). Fortunately, I have very similar results to Cawley. Tables 4.1 and 4.2 present estimates of the OLS coefficients and fixed effect coefficients respectively[1] using the model for log hourly wages, W:

$$LnW_{it} = X_{it}\gamma + \varepsilon_{it} \quad (1)$$

where the subscript i is for the individual, t for time, X_{it} a vector of explanatory variables that include the variable BMI (or another variables measuring weight) that affect wages, and ε_{it} is the residual for observation i in time t (for more details see chapter 3 section 3.6.2).

Columns 1, 4, and 7 of Table 4.1 indicate that, for each ethnic group of females, both BMI and weight in pounds have negative and statistically significant coefficients in OLS regressions. The coefficients in Table 4.1 and the standard

1. For more details see Cawley (2004)

deviations of weight in pounds in Table 4.A1 imply that for two identical white women who differ in weight of two standard deviations (roughly 66 pounds), we would expect the lighter one to enjoy 9.2 percent higher wages, a difference which is roughly equal in magnitude to that associated with 1.5 more years of education, or three more years of work experience. In contrast, for Black females, an increase in weight of two standard deviations (80 pounds) would experience a decrease in wages of 4 percent compared to the median. The same two standard deviations increase for Hispanic females (65 pounds) is associated with a decrease in wages of 5.8 percent.

Columns 1, 4, and 7 of Table 4.2 indicate that the signs and magnitudes of the OLS coefficients on weight vary for males by ethnic group. For white males, the coefficients on BMI and weight are not significantly different from zero. For black males, there is evidence that higher (not lower) body weight is associated with higher wages. Hispanic males suffer a wage penalty from being overweight, and an increase in weight of two standard deviations from the mean weight in pounds is associated with a decrease in wages of seven percent.

Columns 1, 4, and 7 of Tables 4.1 and 4.2 also present the OLS coefficients on the dummy variables for clinical weight classifications (underweight, healthy weight, overweight, and obese where healthy weight is the reference group). Among white females, those who are overweight earn 5 percent less than those of healthy weight, and those who are obese earn 12.5 percent less than those of healthy weight.

Among Black and Hispanic females, the OLS coefficients for the clinical classifications indicate that those who are overweight earn no less than those of healthy weight, while those who are obese earn roughly 5-8 percent less.

In Table 4.2, for white males, the coefficient on the indicator for underweight is negative, that for overweight is positive, and that for obese is negative. For Black males, the OLS coefficients are positive (and significant) on the variables for overweight and for obese. The coefficient for obese Hispanic females is negative and significant at the 1-percent level and on underweight is significant at the 10 percent level.

The weight variables and error term in equation 1 may be correlated if there is reverse causality or if an unobserved variable causes both heaviness and adverse labor market outcomes. One of the approaches Cawley uses to correct for the endogeneity problem is by replacing contemporaneous weight with its 7-year lagged value in the OLS regressions. Columns 2, 5, and 8 in Tables 4.1 and 4.2 present OLS results using a measure of weight lagged seven years, BMI and weight in pounds (rows 1 and 2). For females, the OLS estimates of lagged and contemporaneous measures of the weight variables are similar with all coefficients indicating wage penalties associated with being overweight or obese.

For males, shown in Table 4.2, using contemporaneous weight yielded similar results in most cases with white and Hispanic males subject to wage penalties as weight increases. The two exceptions related to the dummy variable for obese. Being obese 7 years prior is associated with a 9.7 percent reduction in wages compared to white males with a healthy weight, an effect that is double the effect when currently obese is used. The other major difference is with Hispanic males where using currently obese yields a statistically insignificant but negative coefficient while the lagged measure of obesity indicates a statistically significant and negative 8 percent wage penalty compared to those of a healthy weight.

As Cawley (2004) argues, the high degree of similarity between the point estimates on linear measures of weight in the lagged and contemporaneous OLS regressions is consistent with either of two hypotheses: either (1) current wages have little impact on current weight; or (2) current wages do affect current weight, but there is such high serial correlation in both wages and weight that even when distant BMI is used as a regressor, the effect of wages on weight is measured just as strongly. Based on the results in Table 4.2, statement one above may not be true for the exception cases noted above for white males and Hispanic males when the dummy variable formulation of weight is used[2].

Cawley also relies on fixed effect estimators to control for unobservable individual effects. Following Cawley, I estimate a fixed- effects model to eliminate time-invariant heterogeneity. We exploit the panel nature of the NLSY79 data to eliminate individual-specific fixed effects, assuming that the influence of genes and non-genetic factors is constant over time. Columns 3, 6, and 9 of Tables 4.1 and 4.2 report estimates from fixed-effects regressions. The most dramatic difference is that the negative coefficients on BMI and weight in pounds are much smaller (in absolute value) and no longer statistically significant for Black females, Hispanic females, and Hispanic males. Given the just described changes in the results when estimating the fixed effects models, our results suggest that the OLS results for these groups are driven largely by unobserved time-invariant heterogeneity (Similar to the Cawley results).

The coefficients on BMI and weight in pounds are virtually unchanged for white females. The fixed-effects coefficients for Black males are smaller (but the coefficients are positive) than those from OLS and are statistically significant. So far, the finding that heavier white females earn less than normal weight females and overweight Black males earn more than normal weight males is robust.

---

2 .In this chapter, the coefficient on the lagged classification underweight for Black males (-.1701) is greater than in Cawley and is significant at 1-percent level. It is not significant in Cawley's paper. The coefficient in Table 2, column 5 indicates that a Black male who was underweight 7 years earlier earns 17 percent less than a similarly situated healthy weight Black man.

4.2.3 Censored Regressors[3]

Following Rigobon and Stoker (2009), I estimated all the OLS and fixed effect regressions mentioned above with censored regressors (deleting all observations that had missing values instead of creating dummy variables to account for missing variables). Using the same data as Cawley (2004) (NLSY79 from 1981-2000), I compare the results of the two approaches: OLS and fixed effect models including dummy variables for missing observations versus OLS and fixed effect regressions when observations with missing data in the right hand side variables are deleted from the sample.

Tables 4.3 and 4.4 present the empirical evidence for the OLS coefficients and fixed effect coefficients for equation (1) using the data with censored regressors (deleting all the missing values instead of creating dummy variables to account for missing variables).

Comparing the number of observations for females in Table 4.1 and Table 4.3, the number of observations of white women dropped from 26,380 in Table 4.1 to

---

3. Censored regressors refer to the regressors with no missing value: when observations with missing data are deleted from the sample. It was common to drop all observations with any missing values in the past applications in the 1970s. Then dummy variable use became common. Recently Rigobon and Stoker came up with theoretical reasons why the dummy variable approach is questionable.

21,679 in Table 4.3 (4,701 fewer observations). In Table 4.3, we also lost 3,857 observations for Black females and 2,165 for Hispanic females. The smaller sample sizes in Table 4.3 result in higher standard errors. In some cases the coefficients are slightly less significant in Table 4.3 due to the drop in the number of observations.

Comparing the number of observations for males in Tables 4.2 and Table 4.4, the number of observations of white males dropped from 28,754 in Table 4.2 to 23,527 in Table 4.4 (4,701 fewer observations). In Table 4.4, we also lost 4,675 observations for Black females and 2,806 for Hispanic females. The smaller sample sizes in Table 4.4 result in higher standard errors. In some cases the coefficients are slightly less significant in Table 4.4 due to the drop in the number of observations.

Does the method of handling missing data have any effect on the results? Columns 1, 4, and 7 of Table 4.1 and 4.3 indicate the similarity of the results: for each group of females, both the coefficients on BMI and weight in pounds have negative and statistically significant OLS coefficients. The coefficients in Table 4.3 and the standard deviations of weight in pounds for white women imply that for two identical white women who differ in weight by two standard deviations (roughly 66 pounds), we would expect the lighter one to enjoy 9.2 percent higher wages, a difference which is roughly equal in magnitude to that associated with 1.5 more years of education, or three more years of work experience. For Black females, an increase in weight in pounds of two standard deviations (80 pounds) would results in a decrease in wages of

4.8 percent compared to one at the median. The same two standard deviations increase for Hispanic females (67 pounds) is associated with a decrease in wages of 6.7 percent. The penalty is slightly larger for Black and Hispanic females.

Columns 1, 4, and 7 of Table 4.4 indicate that the signs and magnitudes of the OLS coefficients on weight for males vary by ethnic group depending on the way missing data are handled. Here we observe some differences between Table 4.2 and 4.4 in the estimated effects of BMI and weight in pounds for whites and Blacks males. For white males, the coefficients on lagged BMI and lagged weight in pounds are less significant and smaller when compared to results in Table 4.2. In fact, the coefficients are not statistically significant in the censored regressors approach but are statistically significant when dummy variables are used to account for missing observations in the right hand side variables. Another difference associated with using censored regressors data is for Black males where the OLS coefficients on BMI and weight in pounds are not significantly different from zero; they were statistically significant using the Cawley approach.

I used the Wald test to check the hypotheses that the same coefficients (BMI, lagged-7 BMI, weight, and lagged-7 weight) for Black males from the two different regressions (Censored regressors and Cawley's replication) are significantly equal. The hypotheses were rejected, most of the cases, at the 10 percent significance level. The

Wald test confirm that the coefficients of the censored regressors method to handle missing data and Cawley's replication are not equal for Black males

The coefficients on weight for Hispanic males using the censored regressor approach in Table 4.4 are similar to the coefficients for Hispanic males in Table 4.2.

The only other major difference occurred with the fixed effects model where for Black females when employing the dummy variable reflecting being 'overweight', the coefficient on this variable was statistically significant and positive when using the dummy variable approach to handle missing data. However, the coefficient goes to zero when all observations are deleted if they have missing data on the right hand side variables. So my overall conclusion is that it does make a difference which method is used to handle missing data in the right hand side variables.

I also estimated models using dummy variables for clinical weight classification. Columns 1, 4, and 7 of Tables 4.3 and 4.4 present the OLS coefficients on the variables for clinical weight classification. In the censored regressors estimates, there is no evidence for a wage differential between those who are underweight relative to those of healthy weight. Same results were found in OLS regressions for clinical classification used with dummy to account for missing data (Tables 4.1 and 4.2). However white women who are overweight (Table 4.3) earn 4.8 percent less than those of healthy weight, and those who are obese earn 12.5 percent less than those with healthy weight.

Among Black and Hispanic females, the OLS coefficients for clinical weight classification indicate that those who are overweight earn similar wages compared to those of healthy weight, while those who are obese earn roughly 7-9 percent less than those of healthy weight.

As I did when replicating Cawley (2004), I try to eliminate the effect of reverse causality (heavier people may tend to earn less because low wages result in weight gain) by substituting a lagged value of weight for its contemporaneous value. Columns 2, 5, and 8 in Tables 4.3 and 4.4 present OLS results using a measure of weight lagged seven years.

The coefficients on the lagged obese variables for all females as well as white and Hispanic males are negative and statistically significant. White women who were overweight or obese seven years earlier earn 7.9 and 8.81 percent respectively less than women with normal weight. The coefficients are larger and more significant than those found by Cawley. Furthermore, white men, if they are obese seven years earlier, are penalized with 7.7 percent lower wages than a healthy weight white male 7 years earlier. However Hispanic males earned 9.15ercent less if they were obese seven years earlier.

What might be driving the results just discussed associated with the lagged obesity variable? Teenagers with a high BMI, regardless of the reasons, may encounter stigma with regard to their stature and develop poor self-esteem which might negatively affect the accumulation of interpersonal skills or perseverance for social interactions. The lack of such human capital formation may hinder their job performance, and contribute to the BMI wage penalty (Han et al., 2009).

One of the methods used by Cawley to account for the potential endogeneity between the weight variables and the error term in the log wage equation involved estimating a fixed effect model to control for unobservable individual effects. A fixed-effects model is estimated as earlier except this time with censored regressors. Columns 3, 6, and 9 of Tables 4.3 and 4.4 report estimates from fixed-effects regressions. The most dramatic difference with Cawley's results, columns 3, 6 and 9 of Tables 4.1 and 4.2, is that the negative coefficients on BMI and weight in pounds are much smaller and no longer statistically significant for all the race/gender groups except for white female and Hispanic males . This suggests that the OLS results for these groups are driven largely by unobserved time-invariant heterogeneity.

Comparing the coefficients we got from replicating Cawley (Tables 4.1 and 4.2) where he used the dummy variable approach for handling missing data to the censored regressor approach used in Tables 4.3 and 4.4, the coefficients on BMI and weight in pounds are virtually unchanged for white females. The fixed-effects coefficients for Black males are no longer significant. So far, the finding that heavier white females earn less is robust.

4.2.4 Quantile regressions

Quantile regression will help me to see if the coefficients of the weight variables are the same throughout the wage distribution or if they exhibit a pattern. Quantile regressions, which were introduced by Koenker and Bassett (1978), will provide us with the answers. It is particularly useful when the rate of change within the conditional quantile, expressed by the regression coefficients, depends on the quantile being analyzed. I used quantile regressions to investigate the effect of weight on wages across race-ethnic and sex categories dividing the population in five parts: 20th percentile, 40th percentile, 50th percentile, 60th percentile, and 80th percentile where 50th percentile represents the median.

Tables 4.5 and 4.6 report results from the quantile regressions by race and sex, using the data with censored regressors. Table 4.5 indicates that, for each group of females, both BMI and BMI lagged-7 years have negative and statically significant coefficients. As I found before, these effects are largest for white women and smaller for Black and Hispanic females.

Row 2 of Table 4.5 indicates that the wage penalty for white females, using lagged-7 BMI, is 0.86 percent at the first quintile to 1.02 percent at the upper quintile. This result suggests that the least squares estimate of the mean wage penalty, 9.6 percent, is being captured more by the women earning higher wages (40th, 60th, and 80th). The quantile regression delivers more detailed information about the penalty effect of weight on wages. The penalty appears to increase with wages especially for white females and Hispanic females when using the lagged BMI measure. This suggests that the effects of the covariates might not be constant across the distribution of wages. Two possible explanations for these results are that higher paying occupations are more likely to be discriminatory about excessive weight or that women with higher weight seven years earlier are less confident about their physical appearances and less aggressive seeking higher wages. As just noted, for Hispanic women, there is some evidence that the penalty is most severe in the top wage quantile.

Table 5.6 presents a summary of quantile regression results for males. The results for males are again very different from those for females; rows 1 and 2 indicate that only white males of the 20^{th} and 40^{th} quantiles of the distribution earn less if they are heavier. This is a new result not seen with the OLS regression. Rows 2 and 3 of present the estimates for Black males. There is high degree of similarity between the estimates of the OLS regression and the quantile regressions; none of the coefficients are significant, except the coefficient of contemporaneous BMI of the second quantile (40^{TH} percentile) for Black males which is positive and significant at the 5% level (similar to Cawley (2004) but this result is not robust across all regressions). However, for Hispanic males, as seen in rows 5 and 6, the weight penalty increases with the wage brackets. There is similarity between Hispanic females and Hispanic males across all the regressions.

Tables A4.5 and A4.6 report results by race and sex from the quantile regressions, using the same data as Cawley (2004) (The regressors are not censored: dummy variables are created for missing observations). The increase in number of observations results in higher T-statistics and smaller standard errors, so in some cases the coefficients are statistically significant in the quantile regression without censored regressors but not significant in the quantile regression with censored regressors (Tables 4.5 and 4.6)[4]. Table A4.5 row 1 shows the similarity with the censored regressors quantile regression for white females. Row 2 Table A4.5 shows that the weight penalty using lagged BMI is higher for the last two wage brackets and lower for the first two brackets. For Black females, row 4 Table A4.5 shows that the weight penalty is less for the first two quantiles and gets larger for the last two quantiles. For Hispanic females there appears to be an increase in the penalty at higher wages for both BMI and lagged BMI is detected (same results as Hispanic females using the quantile regressions with censored regressors). White males suffer the weight penalty if they are in the 40thor 60th, and 70th quantiles for wage in for both BMI and lagged

4. Although I would like to have controlled for clustering, the STATA software does not allow this option in quantile regressions.

BMI. Finally Hispanic males, for lagged BMI, no more trends but their weight penalty increase as they move to the higher wages brackets 40th, 60th, and 80th.

4.3 Conclusion:

Because of the recent spread of obesity, with its negative economic, social and health consequences, this issue has attracted considerable attention from applied economists. As a result I investigate the relationship between body weight and wages in the U.S. I first replicate Cawley (2004) with the same data (NLSY79 1979-2000), variables, and methods. My results are the same as Cawley's ordinary least squares estimation which indicate that heavier white females, Black females, Hispanic females and Hispanic males tend to earn less, and heavier Black males tend to earn more, than their lighter counterparts. In order to avoid the influence of wages on contemporaneous weight, I estimated models using lagged body weight. Individual fixed effects estimations eliminate the time-invariant unobserved heterogeneous effects on wages. This procedure has a dramatic effect in that it eliminates the negative correlation between BMI and weight in pounds and wages for all groups but White females.

I extend the existing literature by providing new evidence on the relation between weight and wages by asking the question "Does the way missing data is handled cause different results than found by Cawley." His approach utilizes dummy variables for each variable to flag missing data when it occurs. Our approach deletes an observation whenever there is missing data for a particular explanatory variable and is called censoring on the regressors (Rigobon and Stoker (2009)). We are unaware of studies of the effects of weight on wage using this approach for treating missing data. When using this approach, we find that ordinary least squares and individual fixed effects lead to the same results as in Cawley (2004) for all sub-groups but White and Black males. None of the coefficients for white and Black males, contemporaneous weight or 7-years-lagged weight, were statistically significant. I tested the hypotheses that the same coefficients (BMI, lagged-7 BMI, weight, and lagged-7 weight) for Black males from the two different regressions (Censored regressors and Cawley's replication) are significantly equal. The hypotheses were rejected at the 10 percent significance level.

Using Cawley's approach, White males who were obese seven years earlier earn 9.4 percent less than those of normal weight seven years earlier. The coefficients in Table 4.2 and the standard deviations in Table A4.2 imply that for two identical white males who differed such that one was at the mean and one at the two standard deviations above the mean, we would expect the lighter one to enjoy 9 percent higher wages, a difference in magnitude roughly equal to the difference associated with 1.5 more years of education, or three more years of work experience.

I then ran quantile regressions to see to see if the effects of weight on wages are the same throughout the distribution of wages. For instance, are the effects concentrated in the lower part of the wage distribution? I find an increasing weight penalty with wages for almost all sub-groups but for Black males. These findings are important because they are consistent with the results found by Han, Norton, and Powell, (2009) who found that there is indirect effect of BMI in the late teenage years on choices related to both their education and occupation choices. They found that BMI doesn't affect choice of occupation but does affect the wages.

The sociological literature yields a possible explanation for the difference in results (white females are more penalize by weight than Black and Hispanic females) between white females and Black and Hispanic females. Obesity has a more adverse impact on the self-esteem of white females than on that of Black and Hispanic females, who report perceiving higher weight as a signal of power and stability (Stearns 1997). More research is needed to explain differences across gender and race and to explain the increasing penalty associated with higher wages. Is it due to discrimination of employers regarding physical appearance, or is it due to the low self-esteem of the employees with excess weight? Or is it something else?

Table 4.1
Coefficients and t-Statistics from Log Wage Regressions for Females 1981-2000 (Cawley 2004 replication)

	White Female			Black Female			Hispanic Female		
	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects
Column Number	1	2	3	4	5	6	7	8	9
BMI	-0.0086	-0.0089	-0.0069	-0.0034	-0.0051	-0.00009	-0.0054	-0.0066	-0.0009
	(-7.17)	(-5.20)	(-4.25)	(-2.99)	(-3.05)	(-0.05)	(-3.23)	(-3.03)	(-0.37)
Weight in pounds	-0.0014	-0.0015	-0.0011	-0.0005	-0.0008	-0.00008	-0.0009	-0.0011	-0.0002
	(-7.11)	(-5.22)	(-4.25)	(-2.93)	(-3.00)	(-0.30)	(-3.25)	(-3.05)	(-0.49)
Underweight	0.0186	0.0853	-0.0343	-0.0789	-0.0212	-0.0983	-0.0561	-0.0516	0.0148
	(0.91)	(3.1)	(-1.62)	(-2.65)	(-0.67)	(-2.69)	(-1.68)	(-1.16)	(0.41)
Overweight	-0.0507	-0.0733	-0.0358	-0.0078	-0.0251	0.0268	-0.0194	-0.0484	0.0015
	(-4.08)	(-4.19)	(-2.96)	(-0.57)	(-1.37)	(1.93)	(-1.05)	(-1.98)	(0.08)
Obese	-0.1253	-0.0895	-0.096	-0.0524	-0.081	0.0093	-0.081	-0.101	-0.0248
	(-6.75)	(-2.14)	(-4.89)	(-2.86)	(-3.24)	(0.47)	(-3.14)	(-3.20)	(-0.79)
Number of observations	26380	10982	26380	12004	5485	12004	7726	3387	7726

Notes:
1) Data: NLSY79 females
2) One of three measures of weight is used: BMI, weight in pounds (controlling for height in inches) or indicator variables for clinical weight classification: underweight, overweight, and obese (where healthy weight is the excluded categories).
3) For BMI and weight in pounds, coefficients and t-statistics are listed. For indicators of clinical weight classification, the percent change in log wages associated with a change in the indicator variables from 0 to 1 and t statistics are listed.
4) Other regressors include: number of children ever born, age of youngest child, general intelligence, highest grade completed, mother's highest grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, county unemployment rate, current school enrollment, part time job, white collar job, and region of residence.

Table 4.2
Coefficients and t-Statistics from Log Wage Regressions for Males 1981-2000 (Cawley 2004 replication)

	White Male			Black Male			Hispanic Male		
	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects
Column Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
BMI	-0.0015	-0.0039	-0.0003	0.0034	0.0053	0.0038	-0.00588	-0.0076	-0.0003
	(-0.95)	(-1.84)	(-0.18)	(1.92)	(2.03)	(1.56)	(-2.75)	(-2.56)	(-0.15)
Weight in pounds	-0.0002	-0.0006	-0.0002	0.0005	0.0007	0.0004	-0.0009	-0.0011	-0.0005
	(-1.12)	(-2.00)	(-0.95)	(1.98)	(2.1)	(1.34)	(-2.95)	(-2.47)	(-1.24)
Underweight	-0.1544	-0.038	-0.049	-0.0538	-0.1701	0.0582	0.1266	0.1105	0.1206
	(-3.78)	(-0.88)	(-1.34)	(-1.47)	(-2.99)	(1.34)	(1.5)	(1.01)	(1.18)
Overweight	0.0345	0.0118	0.0214	0.0368	0.0093	0.0312	-0.022	-0.0087	0.0022
	(2.7)	(0.71)	(1.89)	(2.35)	(0.45)	(2.09)	(-1.05)	(-0.34)	(0.13)
Obese	-0.0461	-0.0971	-0.0099	0.04054	0.0375	0.0422	-0.04351	-0.08	0.0289
	(-2.33)	(-3.71)	(-0.53)	(1.77)	(1.1)	(1.82)	(-1.52)	(-1.96)	(0.97)
Number of observations	28754	11986	28754	13276	5750	13276	9064	3963	9064

Notes:
1) Data: NLSY79 males
2) One of three measures of weight is used: BMI, weight in pounds (controlling for height in inches) indicator variables for clinical weight classification: underweight, overweight, and obese (where healthy weight is the excluded categories).
3) For BMI and weight in pounds, coefficients and t-statistics are listed. For indicators of clinical weight classification, the percent change in log wages associated with a change in the indicator variables from 0 to 1 and T-statistics are listed.
4) Other regressors include: number of children ever born, age of youngest child, general intelligence, highest grade completed, mother's highest grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, county unemployment rate, current school enrollment, part time job, white collar job, and region of residence.

Table 4.3
Coefficient and t-Statistic from Log Wage Regressions for Females 1981-2000 with Censored Regressors

	White Female			Black Female			Hispanic Female		
	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects
Column Number	1	2	3	4	5	6	7	8	9
BMI	-0.0087	-0.0095	-0.0062	-0.0041	-0.0063	-0.0005	-0.0055	-0.0068	0.0002
	(-6.65)	(-5.18)	(-3.51)	(-2.93)	(-3.08)	(-0.29)	(-3.04)	(-2.79)	-0.07
Weight in pounds	-0.0014	-0.0016	-0.0011	-0.0006	-0.001	-0.0002	-0.0099	-0.0012	-0.0001
	(-6.58)	(-5.17)	(-3.96)	(-2.83)	(-2.99)	(-0.70)	(-3.05)	(-2.86)	(-0.22)
Underweight	0.013	0.0824	-0.0426	-0.0633	0.0324	-0.1145	-0.0365	-0.041	0.0401
	-0.57	-2.75	(-1.77)	(-1.83)	-0.05	(-2.62)	(-0.76)	(-0.72)	-0.94
Overweight	-0.0475	-0.0798	-0.0249	-0.0226	-0.0225	0.0094	0.0188	-0.0521	0.001
	(-3.51)	(-4.32)	(-1.88)	(-1.35)	(-1.05)	-0.59	(-0.88)	(-1.81)	-0.05
Obese	-0.1257	-0.0881	-0.0888	-0.073	-0.0981	-0.0097	-0.0876	-0.0992	-0.0359
	(-6.35)	(-3.18)	(-4.21)	(-3.20)	(-3.11)	(-0.41)	(-3.06)	(-2.78)	(-0.92)
Number of observations	21679	9228	21679	8147	3769	8147	5561	2471	5561

Notes:
1) Data: NLSY79 females
2) One of three measures of weight is used: BMI, weight in pounds (controlling for height in inches) or indicator variables for clinical weight classification: underweight, overweight, and obese (where healthy weight is the excluded categories).
3) For BMI and weight in pounds, coefficients and t-statistics are listed. For indicators of clinical weight classification, the percent change in log wages associated with a change in the indicator variables from 0 to 1 and t-statistics are listed.
4) Other regressors include: number of children ever born, age of youngest child, general intelligence, highest grade completed, mother's highest grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, county unemployment rate, current school enrollment, part time job, white collar job, and region of residence.

Table 4.4
Coeficient and t-Statistic from Log Wage Regressions for Males 1981-2000 with Censored Regressors

	White Male			Black Male			Hispanic Male		
	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects
Column Number	1	2	3	4	5	6	7	8	9
BMI	-0.0006	-0.0023	0.00005	-0.0003	0.0009	0.0036	-0.0066	-0.008	-0.0022
	(-0.36)	(-1.02)	-0.02	(-0.15)	-0.33	-1.2	(-2.64)	(-2.26)	(-0.67)
Weight in pounds	-0.0001	-0.0004	-0.0004	-0.00002	0.0001	0.0002	-0.001	-0.0011	-0.0011
	(-0.52)	(-1.20)	(-1.22)	(-0.08)	-0.35	-0.57	(-2.80)	(-2.19)	(-2.13)
Underweight	-0.1328	-0.0212	-0.0329	-0.0464	-0.123	0.0151	0.1613	0.0472	0.2149
	(-2.75)	(-0.45)	(-0.73)	(-0.92)	(-1.76)	-0.27	-1.42	-0.44	-1.55
Overweight	0.0431	0.0238	0.0253	0.021	-0.0041	0.0411	-0.0286	-0.0087	0.0093
	-3.1	-1.32	-2.06	-1.08	(-0.17)	-2.18	(-1.15)	(-0.29)	-0.47
Obese	-0.0289	-0.0775	0.0036	0.0014	-0.0162	0.0505	-0.0504	-0.0915	0.0358
	(-1.33)	(-2.70)	-0.18	-0.05	(-0.45)	-1.78	(-1.52)	(-1.79)	-0.99
Number of observations	23527	10224	23527	8601	3856	8601	6258	2819	6258

Notes:
1) Data: NLSY79 males
2) One of three measures of weight is used: BMI, weight in pounds (controlling for height in inches) or indicator variables for clinical weight classification: underweight, overweight, and obese (where healthy weight is the excluded categories).
3) For BMI and weight in pounds, coefficients and t-statistics are listed. For indicators of clinical weight classification, the percent change in log wages associated with a change in the indicator variables from 0 to 1 and t statistics are listed.
4) Other regressors include: number of children ever born, age of youngest child, general intelligence, highest grade completed, mother's highest grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, county unemployment rate, current school enrollment, part time job, white collar job, and region of residence.

Table 4.5
Coefficient and t-Statistic from Log Wage Quantile Regressions with Censored Regressors for Females 1981-2000

	20th	40th	50th	60th	80th	# Obs	Row Number
			Quantile Regressions for White Female				
BMI	-0.0081	-0.0081	-0.0078	-0.0081	-0.008	21679	1
	(-8.77)	(-15.94)	(-15.11)	(-14.22)	(-12.73)		
BMI with lagged weight	-0.0086	-0.0095	-0.0094	-0.0093	-0.0102	9228	2
	(-5.98)	(-9.17)	(-9.87)	(-9.8)	(-9.96)		
			Quantile Regressions for Black Females				
BMI	-0.0021	-0.0033	-0.0034	-0.0034	-0.0023	8147	3
	(-3.22)	(-4.86)	(-3.48)	(-3.75)	(-2.92)		
BMI with lagged weight	-0.0041	-0.0051	-0.0047	-0.0048	-0.0051	3769	4
	(-3.55)	(-3.66)	(-2.95)	(-2.92)	(-2.97)		
			Quantile Regressions for Hispanic Females				
BMI	-0.001	-0.0028	-0.004	-0.0038	-0.0063	5561	5
	(-1.12)	(-2.43)	(-4.14)	(-3.51)	(-6.16)		
BMI with lagged weight	-0.0059	-0.0036	-0.0048	-0.0052	-0.0072	2471	6
	(-2.99)	(-2.33)	(-3.04)	(-4.4)	(-3.1)		

Notes:
1) Data: NLSY79 females
2) For BMI and lagged 7 BMI, coefficients and t statistics are listed.
3) Other regressors include: number of children ever born, age of youngest child, general intelligence, highest grade completed, mother's highest grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, county unemployment rate, current school enrollment, part time job, white collar job, and region of residence.

Table 4.6
Coefficients and t-Statistics from Log Wage Quantile Regressions with Censored Regressors for Males 1981-2000

Percentile	20th	40th	50th	60th	80th	# Obs	Row Number
			Quantile Regressions for White Males				
BMI	-0.0014	-0.002	-0.002	-0.0012	0.0004	23527	1
	(-2.31)	(-2.20)	(-2.29)	(-1.40)	(0.39)		
BMI with lagged weight	-0.0014	-0.0041	-0.0039	-0.0029	0.0004	10224	2
	(-0.87)	(-2.47)	(-2.14)	(-1.69)	(0.22)		
			Quantile Regressions for Black Males				
BMI	0.001	0.0028	0.0021	0.0011	0.0017	8601	3
	(0.79)	(3.2)	(2.18)	(1.08)	(0.88)		
BMI with lagged weight	0.0001	0.0014	0.0025	-0.0002	0.003	3856	4
	(0.05)	(0.65)	(0.97)	(-0.895)	(0.68)		
			Quantile Regressions for Hispanic Males				
BMI	-0.0048	-0.0061	-0.0059	-0.005	-0.0068	6258	5
	(-4.82)	(-3.69)	(-3.12)	(-2.49)	(-3.48)		
BMI with lagged weight	-0.0043	-0.0061	-0.0072	-0.0079	-0.0086	2819	6
	(-2.41)	(-2.95)	(-3.21)	(-2.85)	(-3.44)		

Notes:
1) Data: NLSY79 males
2) For BMI and lagged 7 BMI, coefficients and t statistics are listed.
3) Other regressors include: number of children ever born, age of youngest child, general intelligence, highest grade completed, mother's highest grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, county unemployment rate, current school enrollment, part time job, white collar job, and region of residence.

Table A4.1
Summary Statistics for Females NLSY 1981-2000 (uncensored regressors) replicating Cawley's data

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Adjusted wage CPI 2000	46110	11.56	19.23	1	727.9
Log wage	46110	2.24	0.58	0	6.59
Body mass index	46110	25.42	5.86	12.28	83.94
Height in inches corrected	46110	64.24	2.39	51.82	74.44
Weight in pounds corrected	46110	149.32	35.83	76.79	548.99
7-year lag BMI	19854	24.17	5.07	12.48	57.72
7-year lag weight	19854	141.63	31.17	76.79	343.35
General intelligence	46110	40.82	27.81	0	99
Highest grade completed	46110	13.01	2.31	0	20
Mother's highest grade completed	46110	10.43	3.84	0	20
Father's highest grade completed	46110	9.63	5.19	0	20
Year	46110	1990.01	5.59	1981	2000
Unemployment rate	46110	67.56	31.53	0	237
Number of years at current job	46110	3.23	3.87	0	31.76
Years of actual work experience	46110	8.40	5.42	0	23.92
Age	46110	28.93	6.07	16	44
Attending school	46110	0.19	0.39	0	1
Number of kids	46110	1.14	1.25	0	10
Age of youngest child	46110	3.11	4.46	0	27
Hispanic	46110	0.17	0.37	0	1
Black	46110	0.26	0.44	0	1
Underweight	46110	0.04	0.20	0	1
Normal weight	46110	0.54	0.50	0	1
Overweight	46110	0.24	0.42	0	1

Table A4.1
Summary Statistics for Females NLSY 1981-2000 (uncensored regressors) replicating Cawley's data(continued)

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Obese	46110	0.19	0.39	0	1
underw_7	19854	0.05	0.23	0	1
normalw_7	19854	0.62	0.49	0	1
overw_7	19854	0.21	0.40	0	1
obese_7	19854	0.12	0.33	0	1
Work more than 20 hours per week	46110	0.85	0.36	0	1
Work less than 20 hours per week	46110	0.14	0.34	0	1
White collar job	46110	0.62	0.48	0	1
Blue collar job	46110	0.36	0.48	0	1
Married spouse present	46110	0.47	0.50	0	1
Never married	46110	0.36	0.48	0	1
Married spouse not present	46110	0.17	0.38	0	1
Enrolled in school	46110	0.19	0.39	0	1
Not Enrolled in school	46110	0.81	0.39	0	1
Northeast region	46110	0.17	0.38	0	1
North Central region	46110	0.23	0.42	0	1
South region	46110	0.40	0.49	0	1
West region	46110	0.19	0.39	0	1

Table A4.2
Summary Statistics for Males NLSY 1981-2000 (Uncensored regressors) replicating Cawley's data

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Adjusted wage cpi 2000	51101	14.43	23.02	1.00	727.91
Log wage	51101	2.44	0.60	0	6.59
Body mass index	51103	25.83	4.58	10.44	66.72
Height in inches corrected	51103	69.68	2.63	60.53	79.78
Weight in pounds corrected	51103	178.71	34.88	81.94	468.91
7-year lag BMI	21700	24.60	4.13	12.96	54.28
7-year lag weight	21700	170.01	31.85	82.23	412.45
General intelligence	51103	38.83	30.30	0	99
Highest grade completed	51103	12.63	2.47	0	20
Mother's highest grade completed	51103	10.17	4.19	0	20
Father's highest grade completed	51103	9.52	5.30	0	20
year	51103	1990.00	5.53	1981	2000
Unemployment rate	51103	68.09	31.61	0	237
Number of years at current job	51103	3.47	4.08	0	25.92
Years of actual work experience	51103	9.26	5.59	0	23.98
Age	51103	28.81	6.01	16	43
Attending school	51103	0.16	0.36	0	1
Number of kids	51103	0.98	1.23	0	9
Age of youngest child	51103	1.58	3.27	0	32
Hispanic	51103	0.18	0.38	0	1
Black	51103	0.26	0.44	0	1
Underweight	51103	0.01	0.11	0	1
Normal weight	51103	0.48	0.50	0	1

Table A4.2

Summary Statistics for Males NLSY 1981-2000 (uncensored regressors) replicating Cawley's data(continued)

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Overweight	51103	0.35	0.48	0	1
Obese	51103	0.16	0.36	0	1
underw_7	21700	0.02	0.15	0	1
normalw_7	21700	0.59	0.49	0	1
overw_7	21700	0.28	0.45	0	1
obese_7	21700	0.10	0.30	0	1
Work more than 20 hours per week	51103	0.92	0.27	0	1
Work less than 20 hours per week	51103	0.07	0.25	0	1
White collar job	51103	0.35	0.48	0	1
Blue collar job	51103	0.64	0.48	0	1
Maried spouse present	51103	0.44	0.50	0	1
Never married	51103	0.44	0.50	0	1
Married spouse not present	51103	0.12	0.33	0	1
Enrolled in school	51103	0.16	0.36	0	1
Not Enrolled in school	51103	0.84	0.36	0	1
Northeast region	51103	0.18	0.38	0	1
North Central region	51103	0.24	0.43	0	1
South region	51103	0.38	0.48	0	1
West region	51103	0.20	0.40	0	1

Table A4.3
Summary Statistics for Females NLSY 1981-2000 with censored regressors

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Adjusted wage CPI 2000	35387	11.96	19.84	1	727.9
Log wage	35387	2.27	0.58	0	6.59
Body mass index	35634	25.34	5.85	12.28	83.94
Height in inches corrected	35634	64.33	2.39	51.82	74.44
Weight in ounds corrected	35634	149.25	35.69	81.09	548.99
7-year lag BMI	15575	24.07	5.05	14.67	56.83
7-year lag weight	15575	141.42	31.02	81.33	319.52
General intelligence	35634	44.42	27.14	1	99
Highest grade completed	35634	13.2	2.22	0	20
Mother's higest grade completed	35634	11.12	3.04	0	20
Father's highest grade completed	35634	11.09	3.84	0	20
Year	35634	1990.18	5.56	1981	2000
Unemployment rate	35634	69.1	30.63	10	237
Number of years at current job	35634	3.42	3.96	0.02	31.76
Years of actual work experience	35634	8.75	5.43	0.02	23.92
Age	35634	29.08	6.02	16	44
Attending school	35634	0.19	0.4	0	1
Number of kids	35634	1.1	1.23	0	10
Hispanic	35634	0.16	0.36	0	1
Black	35634	0.23	0.42	0	1
Underweight	35634	0.04	0.19	0	1
Normal weight	35634	0.55	0.5	0	1
Overweight	35634	0.23	0.42	0	1
Obese	35634	0.18	0.38	0	1

Table 4.A3
Summary Statistics for Females NLSY 1981-2000 with censored regressors (continued)

Variable	N	Mean	Standard Deviation	Minimum	Maximum
underw_7	15575	0.05	0.22	0	1
normalw_7	15575	0.62	0.48	0	1
overw_7	15575	0.19	0.39	0	1
obese_7	15575	0.12	0.32	0	1
Work more than 20 hours/week	35634	0.86	0.35	0	1
Work less than 20 hours/week	35634	0.14	0.35	0	1
White collar job	35634	0.66	0.47	0	1
Blue collar job	35634	0.34	0.47	0	1
Married spouse present	35634	0.48	0.5	0	1
Never married	35634	0.35	0.48	0	1
Married spouse not present	35634	0.17	0.37	0	1
Enrolled in school	35634	0.19	0.4	0	1
Not Enrolled in school	35634	0.81	0.4	0	1
Northeast region	35634	0.17	0.38	0	1
North Central region	35634	0.24	0.43	0	1
South region	35634	0.4	0.49	0	1
West region	35634	0.19	0.39	0	1
noChild	35634	0.46	0.5	0	1
fiveYChild	35634	0.25	0.43	0	1
tenYChild	35634	0.11	0.31	0	1
eighteenYChild	35634	0.06	0.24	0	1
adultYChild	35634	0	0.05	0	1

Table A4.4
Summary Statistics for Males NLSY 1981-2000 with Censored Regressors

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Adjusted wage cpi 2000	38386	14.97	23.53	1.00	727.91
Log wage	38386	2.48	0.59	0.00	6.59
Body mass index	38612	25.93	4.53	12.33	62.23
Height in inches corrected	38612	69.77	2.60	60.55	79.74
Weight in pounds corrected	38612	179.84	34.62	87.17	468.91
7-year lag BMI	16978	24.65	4.10	13.23	54.28
7-year lag weight	16978	170.87	31.72	86.35	412.45
General intelligence	38612	43.63	29.71	1	99
Highest grade completed	38612	12.89	2.40	3	20
Mother's highest grade completed	38612	11.11	3.17	0	20
Father's highest grade completed	38612	11.10	3.93	0	20
Year	38612	1990.15	5.49	1981	2000
Unemployment rate	38612	69.23	30.50	10	237
Number of years at current job	38612	3.67	4.16	1	25.86
Years of actual work experience	38612	9.56	5.58	0.02	23.98
Age	38612	28.94	5.96	16	43
Attending school	38612	0.16	0.37	0	1
Number of kids	38612	0.96	1.20	0	9
Hispanic	38612	0.16	0.37	0	1
Black	38612	0.22	0.42	0	1
Underweight	38612	0.01	0.10	0	1
Normal weight	38612	0.47	0.50	0	1

Table A4.4
Summary Statistics for Males NLSY 1981-2000 with Censored Regressors (continued)

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Overweight	38612	0.36	0.48	0	1
Obese	38612	0.16	0.37	0	1
underw_7	16978	0.02	0.14	0	1
normalw_7	16978	0.58	0.49	0	1
overw_7	16978	0.29	0.45	0	1
obese_7	16978	0.09	0.29	0	1
Work more than 20 hours per week	38612	0.93	0.25	0	1
Work less than 20 hours per week	38612	0.07	0.25	0	1
White collar job	38612	0.38	0.49	0	1
Blue collar job	38612	0.62	0.49	0	1
Married spouse present	38612	0.45	0.50	0	1
Never married	38612	0.43	0.50	0	1
Married spouse not present	38612	0.12	0.32	0	1
Enrolled in school	38612	0.16	0.37	0	1
Not Enrolled in school	38612	0.84	0.37	0	1
Northeast region	38612	0.18	0.39	0	1
North Central region	38612	0.26	0.44	0	1
South region	38612	0.36	0.48	0	1
West region	38612	0.20	0.40	0	1
noChild	38612	0.61	0.49	0	1
fiveYChild	38612	0.21	0.41	0	1
tenYChild	38612	0.05	0.23	0	1
eighteenYChild	38612	0.02	0.15	0	1
adultYChild	38612	0.00	0.03	0	1

Table A4.5

Coefficients and t-Statistics from Log Wage Quantile Regressions (Uncensored Regressors) for Females 1981-2000

Quantile Regressions for White Female

Percentile	20th	40th	50th	60th	80th	# Obs	Row Number
BMI	-.0075	-.0078	-.0075	-.0080	-.0079	26380	1
	(-14.39)	(-14.65)	(-13.9)	(-18.71)	(-12.99)		
BMI with lagged weight	-.0083	-.0087	-.0082	-.0093	-.0099	10982	2
	(-6.23)	(-15.24)	(-11.38)	(-11.17)	(-11.78)		
Quantile Regressions for Black Females							
BMI	-.0020	-.0027	-.0025	-.0026	-.0021	12004	3
	(-2.51)	(-4.28)	(-4.79)	(-5.54)	(-3.11)		
BMI with lagged weight	-.0032	-.0038	-.0039	-.0047	-.0049	5485	4
	(-2.39)	(-3.65)	(-3.34)	(-3.73)	(-3.28)		
Quantile Regressions for Hispanic Females							
BMI	-.0012	-.0035	-.0039	-.0041	-.0054	7726	5
	(-1.28)	(-4.11)	(-4.43)	(-4.88)	(-3.8)		
BMI with lagged weight	-.0039	-.0034	-.0043	-.0057	-.0092	3387	6
	(-1.71)	(-3.39)	(-3.19)	(-3.05)	(-4.46)		

Notes:

1) Data: NLSY79 females

2) For BMI and lagged-7 BMI, coefficients and t-statistics are listed.

3) Other regressors include: number of children ever born, age of youngest child, general intelligence, highest grade completed, mother's highest grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, county unemployment rate, current school enrollment, part-time job, white collar job, region of residence and dummy variables for missing values.

Table A4.6

Coefficients and t-Statistics from Log Wage Quantile Regressions (Uncensored Regressors) for Males 1981-2000

Percentile	20th	40th	50th	60th	80th	# Obs	Row Number
Quantile Regressions for White Males							
BMI	-.0012	-0.0025	-0.0026	-0.002	-0.0002	28761	1
	(-2.07)	(-2.56)	(-6.35)	(-3.85)	(-0.32)		
BMI with lagged weight	-.0019	-0.0052	-0.0042	-0.0041	-0.0015	11986	2
	(-1.36)	(-3.21)	(-3.14)	(-3.09)	(-0.93)		
Quantile Regressions for Black Males							
BMI	.0024	.0043	.0041	.0052	.0083	13276	3
	(3.51)	(10.56)	(4.45)	(5.63)	(6.31)		
BMI with lagged weight	.0026	.0039	.0050	.0056	.0105	5750	4
	(1.55)	(2.37)	(3.29)	(2.80)	(3.03)		
Quantile Regressions for Hispanic Males							
BMI	-.0046	-.0043	-.0043	-.0043	-.0031	9064	5
	(-2.71)	(-4.28)	(-4.15)	(-4.25)	(-1.97)		
BMI with lagged weight	-.0034	-.0057	-.0066	-.0061	-.0055	3963	6
	(-1.38)	(-2.91)	(-2.77)	(-3.00)	(-2.11)		

Notes:

1) Data: NLSY79 males

2) For BMI and lagged-7 BMI, coefficients and t-statistics are listed.

3) Other regressors include: number of children ever born, age of youngest child, general intelligence, highest grade completed, mother's highest grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, county unemployment rate, current school enrollment, part-time job, white collar job, region of residence and dummy variables for missing values.

Table A4.7

Detailed OLS wage model for males and females using censored regressors

gender=male, race=Hispanic

Linear regression					Number of obs	6258
					F(23, 636)	38.61
					Prob > F	0
					R#NAME?	0.2906
					Root MSE	0.49382

logwage	**Coefficient**	**Robust Std. Err.**	**t-statistic**	**P>|t|**	**[95% Conf.Interval]**	**[95% Conf.Interval]**
Body mass index	-0.0066575	0.0025197	-2.64	0.008	-0.0116053	-0.002
Age	0.014361	0.0066384	2.16	0.031	0.0013252	0.027
Time	-0.0318649	0.0099174	-3.21	0.001	-0.0513397	-0.012
Married spouse present	0.1526438	0.031232	4.89	0	0.0913134	0.214
Married spouse not present	-0.0012356	0.0382061	-0.03	0.974	-0.076261	0.074
Highest grade Completed	0.0360147	0.0074323	4.85	0	0.0214199	0.051
Attending school	-0.1108189	0.0268338	-4.13	0	-0.1635124	-0.058
Highest grade Compl by mother	0.0067368	0.0038444	1.75	0.08	-0.0008125	0.014
Highest grade Compl by father	-5.19E-06	0.0035704	0	0.999	-0.0070164	0.007
General intelligence	0.0027948	0.0007126	3.92	0	0.0013954	0.004
Work less than 20 hours per week	0.031941	0.0558379	0.57	0.568	-0.0777079	0.142

Tenure	0.0169507	0.0031411	5.4	0	0.0107825	0.023119
Years of work experience	0.0252814	0.0056829	4.45	0	0.0141218	0.036441
County unemployment rate	-0.0021749	0.0002718	-8	0	-0.0027086	-0.0016412
White collar	0.0160175	0.0240076	0.67	0.505	-0.0311262	0.0631613
Number of children	0.0097596	0.0115635	0.84	0.399	-0.0129477	0.0324669
Have children 0-5 years	0.0448991	0.022396	2	0.045	0.0009199	0.0888782
Have children 6-10 years	0.0388845	0.0379263	1.03	0.306	-0.0355914	0.1133605
Have children 11-18 years	-0.039902	0.0474081	-0.84	0.4	-0.1329973	0.0531932
Have adult children > 18 years	-0.1488756	0.0739374	-2.01	0.044	-0.2940665	-0.0036847
Northeast region	0.2276743	0.0392732	5.8	0	0.1505534	0.3047951
North Central Region	0.0759851	0.0465577	1.63	0.103	-0.0154404	0.1674105
West region	0.1935869	0.0294457	6.57	0	0.1357643	0.2514094
Constant	1.50926	0.1557744	9.69	0	1.203366	1.815154

gender=male, race=Black

Linear regression					Number of obs	8601
					F(23, 926)	61.98
					Prob > F	0
					R#NAME?	0.3227
					Root MSE	0.45901
logwage	Coefficient	Robust Std. Err.	t-statistic	P>|t|	[95% Conf.Interval]	[95% Conf.Interval]
-------------+-	------------	------------	--------	---------	------------	----------

Body mass index	-0.0003047	0.0020702	-0.15	0.883	-0.0043675	0.0037582
Age	-0.0125743	0.0049272	-2.55	0.011	-0.0222441	-0.0029046
Time	-0.0128488	0.0071568	-1.8	0.073	-0.0268943	0.0011968
Married spouse present	0.1339296	0.0228803	5.85	0	0.0890264	0.1788328
Married spouse not present	0.0842579	0.0316794	2.66	0.008	0.0220861	0.1464297
Highest grade Completed	0.0587862	0.0061763	9.52	0	0.0466651	0.0709072
Attending school	-0.1991542	0.0260562	-7.64	0	-0.2502902	-0.1480182
Highest grade Compl by mother	0.0031632	0.0052654	0.6	0.548	-0.0071703	0.0134967
Highest grade Compl by father	0.0004909	0.003965	0.12	0.902	-0.0072906	0.0082723
General intelligence	0.0038114	0.0005777	6.6	0	0.0026777	0.0049451
Work less than 20 hours per week	0.007453	0.0420115	0.18	0.859	-0.0749957	0.0899017
Tenure	0.0207127	0.0026581	7.79	0	0.0154961	0.0259293
Years of work experience	0.030727	0.0036362	8.45	0	0.0235909	0.0378631
County unemployment rate	-0.0015666	0.0003255	-4.81	0	-0.0022054	-0.0009278
White collar	0.0506054	0.0223349	2.27	0.024	0.0067725	0.0944382
Number of children	-0.0120599	0.0079183	-1.52	0.128	-0.0275999	0.0034801
Have children 0-5 years	0.0422776	0.019644	2.15	0.032	0.0037257	0.0808295
Have children 6-10 years	0.02399	0.0277467	0.86	0.387	-0.0304636	0.0784436
Have children 11-18 years	0.0228521	0.0388139	0.59	0.556	-0.0533214	0.0990256
Have adult children > 18 years	0.1865445	0.1811933	1.03	0.303	-0.1690527	0.5421418
Northeast region	0.1729545	0.0275348	6.28	0	0.1189168	0.2269923
North Central Region	0.0276503	0.0239928	1.15	0.249	-0.0194363	0.0747369
West region	0.2012012	0.0431753	4.66	0	0.1164685	0.2859339
Constant	1.599828	0.1331386	12.02	0	1.33854	1.861116
---------------	------------	------------	--------	---------	------------	-----------

gender=male race=white

Linear regression

Number of obs	23527
F(23, 2549)	159.57
Prob > F	0
R#NAME?	0.3092
Root MSE	0.50199

logwage	Coefficient	Robust Std. Err.	t-statistic	P>|t|	[95% Conf.Interval]	[95% Conf.Interval]
-------------+-	------------	------------	--------	---------	------------	----------
Body mass index	-0.0006523	0.0018361	-0.36	0.722	-0.0042527	0.0029481
Age	-0.0010469	0.0040041	-0.26	0.794	-0.0088986	0.0068047
Time	-0.0137291	0.0058422	-2.35	0.019	-0.0251851	-0.0022732
Married spouse present	0.1241859	0.0165119	7.52	0	0.0918077	0.1565641
Married spouse not present	0.0257728	0.0235139	1.1	0.273	-0.0203356	0.0718811
Highest grade Completed	0.0497906	0.0046096	10.8	0	0.0407517	0.0588294
Attending school	-0.2124943	0.0150267	-14.14	0	-0.2419602	-0.1830284
Highest grade Compl by mother	-0.0030731	0.004114	-0.75	0.455	-0.0111403	0.0049941
Highest grade Compl by father	0.006263	0.003062	2.05	0.041	0.0002588	0.0122672
General intelligence	0.0023175	0.0003666	6.32	0	0.0015986	0.0030364
Work less than 20 hours per week	-0.0976901	0.0255498	-3.82	0	-0.1477905	-0.0475897
Tenure	0.0126265	0.0019834	6.37	0	0.0087372	0.0165157
Years of work experience	0.0275075	0.0034071	8.07	0	0.0208265	0.0341885

County unemployment rate	-0.0012075	0.0002054	-5.88	0	-0.0016103	-0.0008047
White collar	0.0822471	0.014535	5.66	0	0.0537455	0.1107488
Number of children	0.0158882	0.0082521	1.93	0.054	-0.0002933	0.0320696
Have children 0-5 years	0.0339859	0.0131606	2.58	0.01	0.0081794	0.0597924
Have children 6-10 years	-0.0136631	0.0211432	-0.65	0.518	-0.0551227	0.0277966
Have children 11-18 years	-0.0329876	0.0318594	-1.04	0.301	-0.0954605	0.0294854
Have adult children > 18 years	-0.1056524	0.1208289	-0.87	0.382	-0.3425851	0.1312804
Northeast region	0.0832102	0.0205801	4.04	0	0.0428548	0.1235656
North Central Region	-0.0222467	0.0188681	-1.18	0.238	-0.059245	0.0147516
West region	0.0841384	0.0239691	3.51	0	0.0371374	0.1311394
Constant	1.541038	0.0961242	16.03	0	1.352549	1.729528
---------------	------------	------------	--------	---------	------------	----------

gender=female,race=Hispanic

Linear regression					Number of obs	5561
					F(23, 655)	46.24
					Prob > F	0
					R#NAME?	0.3516
					Root MSE	0.45391
logwage	Coefficient	Robust Std. Err.	t-statistic	P>|t|	[95% Conf.Interval]	[95% Conf.Interval]
-------------+-	------------	------------	--------	---------	------------	----------
Body mass index	-0.0055601	0.0018279	-3.04	0.002	-0.0091494	-0.0019709

Age	-0.0104393	0.0064787	-1.61	0.108	-0.0231609	0.0022823
Time	-0.0016583	0.0091206	-0.18	0.856	-0.0195675	0.0162509
Married spouse present	0.0181868	0.0290121	0.63	0.531	-0.0387812	0.0751549
Married spouse not present	0.0169139	0.0329417	0.51	0.608	-0.0477701	0.081598
Highest grade Completed	0.0505097	0.0069549	7.26	0	0.036853	0.0641663
Attending school	-0.1116014	0.0236428	-4.72	0	-0.1580262	-0.0651766
Highest grade Compl by mother	0.0004749	0.0045091	0.11	0.916	-0.0083791	0.0093289
Highest grade Compl by father	0.0018736	0.0039668	0.47	0.637	-0.0059156	0.0096627
General intelligence	0.0028609	0.0007433	3.85	0	0.0014013	0.0043205
Work less than 20 hours per week	-0.004827	0.0316129	-0.15	0.879	-0.0669018	0.0572478
Tenure	0.0162302	0.0039521	4.11	0	0.00847	0.0239905
Years of work experience	0.0358887	0.0041394	8.67	0	0.0277606	0.0440169
County unemployment rate	-0.0012607	0.0002705	-4.66	0	-0.0017919	-0.0007295
White collar	0.1009831	0.0226787	4.45	0	0.0564514	0.1455149
Number of children	0.0162893	0.0126176	1.29	0.197	-0.0084865	0.0410651
Have children 0-5 years	-0.0260772	0.0212951	-1.22	0.221	-0.0678921	0.0157378
Have children 6-10 years	0.0310349	0.0255629	1.21	0.225	-0.0191602	0.0812299
Have children 11-18 years	0.0233562	0.0348873	0.67	0.503	-0.0451482	0.0918606
Have adult children > 18 years	-0.0162953	0.1136455	-0.14	0.886	-0.2394488	0.2068581
Northeast region	0.2070903	0.0367681	5.63	0	0.1348927	0.2792879
North Central Region	-0.0321324	0.043133	-0.74	0.457	-0.1168279	0.0525632
West region	0.0750617	0.0278957	2.69	0.007	0.0202859	0.1298375
Constant	1.583468	0.1477196	10.72	0	1.293407	1.873529
---------------	------------	------------	--------	---------	------------	----------

gender=female,race=Black

Linear regression					Number of obs	8147
					F(23, 964)	75.93
					Prob > F	0
					R#NAME?	0.3648
					Root MSE	0.42282
logwage	Coefficient	Robust Std. Err.	t-statistic	P>|t|	[95% Conf.Interval]	[95% Conf.Interval]
-------------+-	------------	------------	--------	---------	------------	----------
Body mass index	-0.0041393	0.001414	-2.93	0.003	-0.0069143	-0.0013644
Age	-0.0108988	0.00437	-2.49	0.013	-0.0194746	-0.002323
Time	0.0077641	0.0065469	1.19	0.236	-0.0050837	0.0206119
Married spouse present	-0.0053285	0.0194445	-0.27	0.784	-0.043487	0.03283
Married spouse not present	0.0423351	0.0257882	1.64	0.101	-0.0082724	0.0929425
Highest grade Completed	0.0468472	0.0061547	7.61	0	0.0347691	0.0589253
Attending school	-0.109458	0.0184843	-5.92	0	-0.1457322	-0.0731838
Highest grade Compl by mother	0.0020808	0.0045133	0.46	0.645	-0.0067762	0.0109378
Highest grade Compl by father	-0.0010179	0.0033024	-0.31	0.758	-0.0074987	0.0054628
General intelligence	0.0041887	0.0005865	7.14	0	0.0030377	0.0053397
Work less than 20 hours per week	-0.0420098	0.0222591	-1.89	0.059	-0.0856916	0.001672
Tenure	0.0182208	0.0026618	6.85	0	0.0129972	0.0234444
Years of work experience	0.0307747	0.0030835	9.98	0	0.0247235	0.0368259
County unemployment rate	-0.000393	0.0003064	-1.28	0.2	-0.0009942	0.0002082
White collar	0.0985653	0.0185336	5.32	0	0.0621945	0.1349362

Number of children	-0.0153746	0.0090219	-1.7	0.089	-0.0330795	0.0023303
Have children 0-5 years	0.026287	0.01484	1.77	0.077	-0.0028354	0.0554094
Have children 6-10 years	0.0216086	0.0202098	1.07	0.285	-0.0180517	0.0612689
Have children 11-18 years	0.0099687	0.0258191	0.39	0.7	-0.0406994	0.0606367
Have adult children > 18 years	-0.0334627	0.051984	-0.64	0.52	-0.1354775	0.0685522
Northeast region	0.2197096	0.0259269	8.47	0	0.16883	0.2705892
North Central Region	0.0736369	0.0255776	2.88	0.004	0.0234426	0.1238312
West region	0.1376989	0.0382375	3.6	0	0.0626606	0.2127372
Constant	1.468212	0.1183069	12.41	0	1.236043	1.70038
---------------	------------	------------	--------	---------	------------	----------

<u>gender=female,race=white</u>

Linear regression

Number of obs	21679
F(23, 2693)	147.32
Prob > F	0
R#NAME?	0.3495
Root MSE	0.48779

logwage	Coefficient	Robust Std. Err.	t-statistic	P>|t|	[95% Conf.Interval]	[95% Conf.Interval]
-------------+-	------------	------------	--------	---------	------------	----------
Body mass index	-0.0087114	0.0013097	-6.65	0	-0.0112796	-0.0061433
Age	-0.0039461	0.0035441	-1.11	0.266	-0.0108956	0.0030034
Time	-0.0045639	0.0051535	-0.89	0.376	-0.0146691	0.0055413

Married spouse present	0.0311537	0.0144842	2.15	0.032	0.0027524	0.0595549
Married spouse not present	0.0588852	0.0195205	3.02	0.003	0.0206085	0.0971619
Highest grade Completed	0.0625932	0.0043196	14.49	0	0.0541232	0.0710632
Attending school	-0.1036961	0.0126006	-8.23	0	-0.1284039	-0.0789883
Highest grade Compl by mother	0.000581	0.0035691	0.16	0.871	-0.0064176	0.0075795
Highest grade Compl by father	0.00417	0.0026633	1.57	0.118	-0.0010524	0.0093923
General intelligence	0.0020262	0.0003396	5.97	0	0.0013602	0.0026921
Work less than 20 hours per week	-0.0605749	0.0163223	-3.71	0	-0.0925804	-0.0285694
Tenure	0.0211685	0.0019088	11.09	0	0.0174256	0.0249113
Years of work experience	0.0317596	0.0028686	11.07	0	0.0261347	0.0373844
County unemployment rate	-0.0010626	0.000194	-5.48	0	-0.0014431	-0.0006822
White collar	0.1659594	0.0138098	12.02	0	0.1388805	0.1930383
Number of children	-0.0294661	0.0082874	-3.56	0	-0.0457164	-0.0132158
Have children 0-5 years	-0.0188596	0.0134157	-1.41	0.16	-0.0451657	0.0074465
Have children 6-10 years	-0.057396	0.0174246	-3.29	0.001	-0.091563	-0.023229
Have children 11-18 years	-0.0838483	0.0240957	-3.48	0.001	-0.1310962	-0.0366004
Have adult children > 18 years	-0.0466925	0.063481	-0.74	0.462	-0.1711689	0.077784
Northeast region	0.102731	0.0191546	5.36	0	0.0651718	0.1402902
North Central Region	-0.0219592	0.0169751	-1.29	0.196	-0.0552447	0.0113262
West region	0.1038644	0.020941	4.96	0	0.0628022	0.1449265
Constant	1.270984	0.0892083	14.25	0	1.096061	1.445908

Chapter 5

EFFECTS OF OBESITY ON WAGES: 1981-2006

In chapter 5, I continue to examine the wage penalty associated with being overweight or obese. We extend the results from Chapter 4 in two ways. First, the sample is expanded to cover the years 2002, 2004 and 2006 which to my knowledge have not been used in wage penalty research. Second, although we found little difference in the results using the Cawley's approach to missing data and the approach when observations are deleted if there are missing data on a regressor, the second approach is chosen. Finally, we look at the effects of obesity when using quantile regression estimates.

The outline of this chapter is as follows. Section 5.1 is an introduction while 5.2 presents the empirical results, where section 5.2.1 presents the descriptive statistics for the two samples underlying our results, section 5.2.2 discusses the results of the analysis using data with censored regressors and 5.2.3 presents the quantile regressions with censored regressors. I conclude in section 5.3.

5.1 Introduction

This chapter explores the effects of weight on wages and compares the results with the ones from chapter 4. I use a larger data set National Longitudinal Survey of Youth (NLSY 1979-2006). The data extends three years (2002, 2004, and 2006) more than the data used in chapter 4 (NLSY 1979-2000). Nevertheless this chapter used several regression strategies, in an attempt to generate consistent estimates of the effect of weight on wages and to identify any change in the penalty across different wage brackets.

Like our results in Chapter 4, I attempt to confront the fact that a correlation between BMI and wages needs to be a causal relationship running from the BMI (weight) to the wages. We use the procedures used in Chapter 4 to account for the potential correlation between the weight variables and the error term in the log wage equation. This is done by replacing the contemporaneous weight with its 7-years lagged value[1] or with its 8-years lagged value if the 7-years lagged variable does not exist. Fixed effects estimators are also used to control for unobserved individual effects.

In this chapter the results are based solely on observations with complete data on the

1. I used the same lagged 7-year value as the literature to able to compare results

regressors. Quantile regressions investigate whether the effects of weight vary across the wage distribution. The results in chapter 5 find that excessive weight lowers wages for white females. OLS estimates for white women indicate that a difference in weight of two standard deviations (roughly 68 pounds) is associated with a difference in wage of 9.5 percent. In absolute value, this is equivalent to the effects of roughly two years of education. Negative correlations between weight and wages observed for other gender-ethnic groups appear to be due to unobserved heterogeneity. Also I find some evidence that the weight penalty for some gender-ethnic groups gets larger as wages increase.

The starting point of my empirical study starts by running OLS and fixed effects models using censored regressors procedure to test the robustness of the previous results across different gender and race groups. I conclude by using quantile regressions, in order to see if the coefficients vary over the wage distribution.

5.2 <u>Empirical Results</u>

5.2.1 Descriptive Statistics:

Appendices 5.A1, 5.A2, 5.A3, and 5.A4 provide summary statistics by gender. Table 5.A1 and 5.A2 provide descriptive statistics of the samples of females and males respectively, replicating Cawley 2004. That is, the sample includes persons who have missing data in the regressors.

Tables 5.A3 and 5.A4 present the descriptive statistics for the samples of females and males respectively after observations are deleted that have any missing data for a right-hand side variable. Next, we compare the means for some key variables by gender. The mean age for all the samples is age 31.

First we will examine means for females. For the most part the means are fairly close for both samples. In both tables the mean BMI is around 26, height is 64 inches, and weight is just less than 153 pounds. The exceptions are for wages, mother's years of education, father's years of education, test of general intelligence, and age of the youngest child. The mean wage (2006 dollars) in the censored sample is $14.52 while in the sample including observations with missing data the mean is $11.09. Means for mother's highest grade completed, father's highest grade completed, general intelligence and age of youngest child were for the Cawley sample (censored regressors sample) respectively 10.40 (11.09), 9.57 (11.05), 40.50 (44.21) and 2.06 (2.09).

As with the females, the means for males are fairly close in both samples. In both tables the mean BMI is around 26.3, height is 69.7 inches, and weight is about 182 pounds. The exceptions are for wages, mother's years of education, father's years of education, test of general intelligence, and age of the youngest child. The mean wage (2006 dollars) in the censored sample is $18.44 while in the sample including observations with missing data the mean is $17.86 Means for mother's highest grade completed, father's highest grade completed, and general intelligence were for the Cawley sample (censored regressors sample) respectively 10.08 (11.11), 9.53 (11.10), and 38.94 (43.78).

5.2.2 Censored Regressors[2]

Tables 5.1 and 5.2 present OLS coefficient estimates for the weight related regressors indicating contemporaneous weight in pounds, OLS estimates with lagged weights in pounds, and fixed effects estimates for equation (3.1). Table 5.1 includes the results for females by ethnicity and Table 5.2 includes results for males by ethnicity. I estimated the OLS and fixed effects ln wage regressions using the National Longitudinal Survey of Youth NLSY1979-2006 with censored regressors (following Rigobon and Stoker (2009)).

---

2. Censored regressors refer to the regressors with no missing value: observations with missing data are deleted from the sample.

Recall the model for log hourly wages, W, for individual i in time t is:

$$LnW_{it} = X_{it}\gamma + \varepsilon_{it} \tag{3.1}$$

where X_{it} is a vector of explanatory variables that includes variable BMI (or another variables measuring weight), and ε_{it} is the residual for observation i in time t. The regressors are the same as were used in Chapter 4 and are defined in section 3.3. The complete list of variables that are included in addition to the weight variables are shown below the tables.

Our results for females in columns 1, 4, and 7 of Table 5.1 indicate that, for each group, both contemporaneous BMI and weight in pounds have negative and statistically significant coefficients in OLS regressions. The coefficients in Table 5.1 and the means and standard deviations of weight in pounds in Table A1 imply that for two identical white women who differed in weight by two standard deviations (roughly 68 pounds), we would expect the lighter one to enjoy 9.5 percent higher wages. This wage difference is roughly equal in magnitude to that associated with two more years of education, or four more years of work experience. For Black females, an increase in weight (in pounds) of two standard deviations (80 pounds) from the mean weight is associated with a decrease in wages of 6.4 percent. The same two standard deviations increase for Hispanic females (70 pounds) result in a decrease in wages of 7.7 percent.

The results for males are shown in Table 5.2 where columns 1, 4, and 7 indicate that the signs and magnitudes of the OLS coefficients on weight for males vary by ethnic group. For white males, the coefficients on BMI and weight in pounds are smaller and not as significant in the statistical sense when compared to results in Table 4.3 (chapter 4) which include data on up to 2000. In other words, the effects of wage appear to have gotten smaller by including the three years of data from the 21st century. The OLS coefficients on BMI and weight in pounds for Black males are not significantly different from zero when three years of the 21st century are included in the data. The coefficients on weight for Hispanic males are less significant than the coefficients for Hispanic males in Table 4.3 (chapter 4). These results are likely to be due to adding older persons to our sample (age 42 -49). One might speculate that a Hispanic male who is 49 years old is less likely to be affected by his weight than a male in his early carrier.

I also estimated models including dummy variables for clinical weight classifications Underweight, Overweight, and Obese (healthy weight is the omitted group). Columns 1, 4, and 7 of Tables 5.1 and 5.2 present the OLS coefficients on the indicator variables for clinical weight classifications for females and males repectively. In the censored regressors, there is no evidence or a wage differential between those who are underweight relative to those of healthy weight. However white women who are overweight earn 3.8 percent less than those classified as having a healthy weight, and those who are obese earn 13.1 percent less than those having a healthy weight.

Among black and Hispanic females, the OLS coefficients for clinical weight classification indicate that those who are overweight earn no less than those having a healthy weight, while those who are obese earn roughly 7-10 percent less than those of healthy weight. The coefficients on underweight, overweight, and obese are not significant for Hispanic males.

In Table 5.2, for white males, the coefficient on the dummy variable indicating that a person is underweight is negative, and the coefficient for overweight is positive. If we take the beauty norms of our century, the probability of being considered attractive for a male with a BMI less than 20, is small. Consequently, due to client discrimination or self-esteem, he might end up with lower wages comparing to a healthy weight male, all other variables considered constant. For Black males, the OLS coefficient on the dummy variable for overweight is positive and significant at the 10 percent level.

Endogeneity might arise if there is reverse causality or if an unobserved variable causes both heaviness and adverse labor market outcomes. Cawley solves the endogeneity problem by replacing contemporaneous weight with its 7-year lagged value (8-year if 7-year is not available) in the OLS regression[3]. He also relies on fixed effect estimators to control for unobservable individual effects.

Columns 2, 5, and 8 in Table 5.1for females and Table 5.2 for males present OLS results using a measure of weight lagged seven years. The OLS estimates of lagged and contemporaneous BMI and weight are similar even though the lagged weight variables result in smaller sample sizes compared to the sample including contemporaneous weight variables As Cawley (2004) argues, the high degree of similarity between the point estimates on linear measures of weight in the lagged and contemporaneous OLS regressions is consistent with either of two hypotheses: either (1) current wages have little impact on current weight; or (2) current wages do affect current weight, but there is such high serial correlation in both wages and weight that even when distant BMI is used as a regressor, the effect of wages on weight is measured just as strongly.

In Table 5.2, the coefficient on the dummy variable indicating being underweight 7 years earlier, is negative and statistically significant for Black males

3. The data available are sixteen years (1981, 1982, 1985, 1986, 1988, 1989, 1990, 1992, 1993, 1994, 1996, 1998, 2000, 2002, 2004, and 2006). For the years 1988, 1989, 1992, 1993, 1996, 1998, and 2000 I used lagged 7- years. For the years where the measure of weight lagged 7-years is not available, we used lagged 8-years if available. These years are 1990, 1994, 1998, 2002, 2004, and 2006.

(-.1657) and is consistent with my finding in chapter 4. The coefficient in Table 5.2 column 5 indicates that a Black male who was underweight 7 years earlier earns 16 percent less than a Black man having a healthy weight. "Teenagers with high BMI, regardless of the reasons, may encounter stigma with regard to their stature and develop poor self-esteem which might affect the accumulation of interpersonal skills or perseverance for social interactions. The lack of such human capital formation may hinder their performance in their jobs, and contribute to the BMI wage penalty" (Han, Euna Norton, Edward C. Powell, Lisa M. 2009). This hypothesis is consistent with our results. In column 2 of Table 5.1: underweight white women seven years earlier earn 6.3 percent more than females having a normal weight. Also white women who were overweight or obese seven years earlier earn 7 and 10 percent respectively less than women with normal weight. The coefficients are larger and more significant than those found by Cawley. Furthermore, white men, if they were obese early in their careers, are penalized with 7 percent lower wages. For Black males being underweight 7 or 8 years prior is associated with wages that are lower by 16 percent when compared to those who had a healthy weight 7 or 8 years earlier. No other significant effects are found for Black males using lagged weight. However Hispanic males earned 9.8 percent less than a healthy weight Hispanic male if they were obese seven years earlier.

A fixed- effects model is estimated to eliminate time-invariant heterogeneity. We exploit the longitudinal nature of the NLSY79 data to eliminate individual-specific fixed effects, assuming that the influence of genes and non-genetic factors is constant over time.

Tables 5.1 and 5.2, columns 3, 6, and 9 report results from fixed-effects regressions for females and males respectively. The most dramatic difference with the weight and lagged weight coefficients is that the negative coefficients on BMI and weight in pounds are much smaller and no longer statistically significant for Black females, Hispanic females, and Hispanic males. This suggests that the OLS results for these groups are driven largely by unobserved time-invariant heterogeneity (as was the case with Cawley's results). Nevertheless, columns 3, and 6, of Table 5.2 show some evidence that overweight white males, overweight Black males and obese Black males enjoy premiums of 3 percent, 5 percent, and 2 percent respectively more compared to healthy white and Black males.

Comparing my results with those of Cawley (2004), the coefficients on BMI and weight in pounds are virtually unchanged for white females. So far, the finding that heavier white females earn less is robust through all the models and in the data up to 2000 and the data set which includes 2002, 2004, and 2006.

5.2.3 Quantile regressions

Are the effects of weight on wages the same throughout the distribution of wages? For instance, are the effects concentrated in the lower part of the wage distribution?

Quantile regressions, which were introduced by Koenker and Bassett (1978), will provide us with the answers. It is particularly useful when the rate of change within the conditional quantile, expressed by the regression coefficients, depends on the quantile being analyzed. I used quantile regressions to investigate the effect of weight on wages across race-ethnic and sex categories dividing the population in five parts: 20^{th} percentile, 40^{th} percentile, 50^{th} percentile, 60^{th} percentile, and 80^{th} percentile where 50^{th} percentile represents the median.

Tables 5.3 and 5.4 report results from the quantile regressions for females and males respectively, using the censored regressors data from the NLSY79 for the years (1979-2006). Table 5.3 indicates that, for each group of females, both BMI and lagged BMI have negative and statistically significant coefficients. As in all my previous results, these effects are largest for white women and smaller for Black and Hispanic females. Row 2 of Table 5.3 indicates that the wage penalties for white females, using lagged BMI, are 0.81 percent at the 20^{th} percentile and 1.03, 1.02 and 1.09 percent at the 40^{th}, 60^{th}, and 80^{th} percentile. These results suggest that the least squares estimate of the wage penalty is being borne more by the women earning higher wages. The quantile regression delivers more detailed information about the effect of weight on wages. The penalty appears to increase with wages, suggesting that the effects of the covariates are not constant along the wage distribution. Consequently, lower wage white women workers enjoy smaller penalty. These could be due to the appearance of low skilled workers is not as important as for high skill occupations. Two possible explanations for these results are possible. Higher paying occupations may be more likely to be discriminatory about excessive weight. Alternatively women with higher weight seven years earlier might be less confident about their physical appearances and less aggressive seeking higher wages.

Rows 4 and 6 of Table 5.3 indicate that, among Black and Hispanic females, those who were overweight seven years earlier, earn less than those having a healthy weight for all five brackets of wage. Nevertheless, the 40th, 50th and 60th percentile suffer slightly greater penalty for Black females than the 20th percentile. However, we can depict an increasing penalty for Hispanic females: the wage penalties are most severe in the top wage quantile (60th and 80th)for both BMI and lagged BMI.

Table 5.4 presents a summary of quantile regression results for males. The results for males are very different from those for females. Row 2 of Table 5.4 indicates that only white males of the middle three quantiles, 40th, 50th, and 60th, have negative and significant at the 10, 10 and 5 percent level respectively. This is a new result not seen with the OLS regression.

Rows 2 and 3 of Table 5.4 present the estimates for Black males. There is high degree of similarity between the estimates of the OLS regression and the quantile regressions. None of the coefficients of lagged BMI are significant. Nevertheless, the coefficients of contemporaneous BMI for Black males are positive and significant at the 5% level (similar to Cawley (2004) but this result is not robust across all regressions. However, for Hispanic males row 6 of Table 5.4 shows an increasing penalty with wage if we compare the first two quantiles to the last two quantiles. The wage penalty rises as the weight increases for Hispanic males.

Tables 5.A7 and 5.A8 report the estimated coefficients and t-statistics by race and gender resulting from the quantile regressions, using the NLSY79 (1979-2006) uncensored (dummy variables are created to account for missing observations). We got the same results found in chapter 4 (for more details see section 4.2.4)[4].

---

4. The main problem with the quantile regression: I was not able to control for clustering (due to the limited capabilities of the software).

5.3 Conclusion:

This chapter investigates and measures the correlation between body weight and wages in the US. We add 3 years of data to the sample size in Chapter 4, and delete the observations with missing values (censored regressors) in the right hand side variables. Ordinary least squares results indicate that heavier white females, Black females, and Hispanic females tend to earn less than their lighter counterparts. Models also are estimated using lagged body weight, in order to account for the possible influence of wages on contemporaneous weight. Results from these regressions are consistent with wages having little effect on contemporaneous weight. White males who were obese seven years earlier earn 7 percent less than those of normal weight seven years earlier.

Fixed effects models allow us to take advantage of the panel nature of the NLSY79 data. These fixed effects models allow us to account for the influence of time invariant unobserved heterogeneity on weight and wages. The fixed effect model has a dramatic effect. It eliminates the negative correlation between weight and wages for all but for white females.

Finally, quantile wage regressions were then estimated to see if the effects of weight on wages are the same throughout the distribution of wages. For instance, are the effects concentrated in the lower part of the wage distribution? Or it is concentrated in the upper part? Or is it concentrated in the middle? I find an increasing weight penalty with wages some sub-groups.

The sociological literature yields a possible explanation for the difference in results between white females and Black and Hispanic females. White females are more penalized by weight than Black and Hispanic females. Obesity has a more adverse impact on the self-esteem of white females than on that of Black and Hispanic females, who report perceiving higher weight as a signal of power and stability (Stearns 1997). More research is needed to explain differences across gender and race and to explain the increasing penalty associated with higher wages. For example, is it due to employers of workers who directly interact with customers may simply be passing on customers' prejudice against overweight employees. Or is it due to the low self-esteem of the employees with excess weight? Or is it something else?

Table 5.1
Coefficient and t-Statistic from Log Wage Regressions for Females 1981-2006 with Censored Regressors

	White Female			Black Female			Hispanic Female		
	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects
Column Number	1	2	3	4	5	6	7	8	9
BMI	-.0086	-.0099	-.0046	-.0043	-.0071	0.0026	-.0065	-.0085	-0.0008
	(-6.66)	(-5.66)	(-2.66)	(-2.96)	(-3.75)	(1.23)	(-3.44)	(-3.49)	(-0.28)
Weight in pounds	-.0014	-.0016	-.0008	-.0008	-.0011	.0002	-.0011	-.0015	-.0003
	(-6.60)	(-5.62)	(-3.06)	(-3.05)	(-3.71)	(0.77)	(-3.47)	(-3.55)	(-0.58)
Underweight	-0.0007	.0639	-0.0441	-.08040	-.0292	-0.1480	-.0359	-.0646	0.3094
	(-0.03)	(2.31)	(-1.80)	(-1.91)	(-0.72)	(-3.29)	(-0.76)	(-1.19)	(0.69)
Overweight	-.0384	-.0780	-0.0108	-.0265	-.0272	.0225	-0.0271	-.0529	0.0002
	(-2.75)	(-4.43)	(-0.77)	(-1.53)	(-1.38)	(1.32)	(-1.18)	(-1.87)	(0.01)
Obese	-.1314	-.1028	-0.0807	-.0797	-.1087	0.0230	-.0998	-.1166	-.0385
	(-6.56)	(-3.97)	(-3.70)	(-3.44)	(-3.82)	(0.89)	(-3.28)	(-3.24)	(-1.05)
Number of observations	24099	15889	21679	9320	6623	9320	6300	4356	6300

Notes:
1) Data: NLSY79 females
2) One of three measures of weight is used: BMI, weight in pounds (controlling for height in inches) or indicator variables for clinical weight classification: underweight, overweight, and obese (where healthy weight is the excluded categories).
3) For BMI and weight in pounds, coefficients and t-statistics are listed. For indicators of clinical weight classification, the percent change in log wages associated with a change in the indicator variables from 0 to 1 and t-statistics are listed.
4) Other regressors include: number of children ever born, age of youngest child, general intelligence, highest grade completed, mother's highest grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, county unemployment rate, current school enrollment, part time job, white collar job, and region of residence.

Table 5. 2
Coeficient and t-Statistic from Log Wage Regressions for Males 1981-2006 with Censored Regressors

	White Male			Black Male			Hispanic Male		
	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects
Column Number	1	2	3	4	5	6	7	8	9
BMI	-.0003	-.0017	.0012	.00104	.0013	.0040	-.0064	-.0081	0.0010
	(-0.17)	-0.78	(0.57)	0.48	0.53	(1.32)	(-2.62)	-2.52	(0.33)
Weight in pounds	-.00009	-.0003	-.00002	.00016	.00018	.0005	-.0010	-.0012	-.0001
	(-0.36)	-0.97	(-0.07)	(0.55)	0.50	(1.19)	(-2.75)	-2.56	(-0.36)
Underweight	-.1303	-.0142	-0.0352	-.0447	-.1657	.0186	.1603	-.0332	0.2164
	(-2.87)	-0.28	(-0.78)	(-0.94)	-2.69	(0.36)	(1.35)	-0.33	(1.54)
Overweight	.0509	.0296	0.0368	.0369	.0197	0.0526	-.0204	-.0131	0.0232
	(3.57)	1.76	(3.05)	(1.91)	0.94	(2.53)	(-0.81)	-0.46	(1.08)
Obese	-.0203	-.0691	0.020	0.0198	-.0225	0.0289	-.0494	-.0981	0.0581
	(-0.95)	-2.60	(1.03)	(-0.76)	-0.68	(2.14)	(-1.48)	-2.21	(1.58)
Number of observations	25951	17566	25951	8601	6728	8603	7019	4891	7021

Notes:
1) Data: NLSY79 males
2) One of three measures of weight is used: BMI, weight in pounds (controlling for height in inches) or indicator variables for clinical weight classification: underweight, overweight, and obese (where healthy weight is the excluded categories).
3) For BMI and weight in pounds, coefficients and t-statistics are listed. For indicators of clinical weight classification, the percent change in log wages associated with a change in the indicator variables from 0 to 1 and t-statistics are listed.
4) Other regressors include: number of children ever born, age of youngest child, general intelligence, highest grade completed, mother's highest grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, county unemployment rate, current school enrollment, part time job, white collar job, and region of residence.

Table 5.3

Coefficients and t-Statistics from Log Wage Quantile Regressions with Censored Regressors for Females 1981-2006

Percentile	20th	40th	50th	60th	80th	# Obs	Row Number
			Quantile Regressions for White Female				
BMI	-0.0074	-0.008	-0.008	-0.008	-0.0084	21679	1
	(-10.66)	(-14.5)	(-14.25)	(-14.4)	(-11.74)		
BMI with lagged weight	-0.0081	-0.0103	-0.0102	-0.0109	-0.0097	9228	2
	(-7.25)	(-11.34)	(-12.56)	(-14.3)	(-9.34)		
			Quantile Regressions for Black Females				
BMI	-0.003	-0.0028	-0.0038	-0.0044	-0.0029	8147	3
	(-3.78)	(-4.35)	(-4.57)	(-4.25)	(-2.78)		
BMI with lagged weight	-0.0039	-0.0052	-0.0054	-0.0062	-0.0046	3769	4
	(-4.57)	(-4.39)	(-6.57)	(-6.48)	(-5.12)		
			Quantile Regressions for Hispanic Females				
BMI	-0.0034	-0.003	-0.0048	-0.0049	-0.007	5561	5
	(-2.32)	(-2.84)	(-4.6)	(-4.36)	(-6.58)		
BMI with lagged weight	-0.0055	-0.0049	-0.0052	-0.0071	-0.01	2471	6
	(-4.26)	(-4.31)	(-4.19)	(-6.1)	(-7.35)		

Notes:
1) Data: NLSY79 females
2) For BMI and lagged 7 BMI, coefficients and t statistics are listed.
3) Other regressors include: number of children ever born, age of youngest child, general intelligence, highest grade completed, mother's highest Grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, county unemployment rate, current school enrollment, part time job, white collar job, and region of residence.

Table 5.4
Coefficients and t-Statistics from Log Wage Quantile Regressions with Censored Regressors for Males 1981-2006

Percentile	20th	40th	50th	60th	80th	# Obs	Row Number
			Quantile Regressions for White Males				
BMI	0.0008	-.0018	-.0009	-.0015	.0011	23527	1
	(-1.32)	(-1.88)	(-1.10)	(-1.73)	(1.09)		
BMI with lagged weight	-.0014	-.0019	-.0018	-.0021	.0014	10224	2
	(-1.45)	(-1.65)	(-1.81)	(-2.10)	(1.21)		
			Quantile Regressions for Black Males				
BMI	0.0018	.0030	.0027	.0027	.0035	8601	3
	(1.29)	(3.34)	(3.32)	(3.37)	(3.20)		
BMI with lagged weight	.0039	.0018	.0017	.0012	.0022	3856	4
	(1.56)	(0.96)	(1.16)	(0.75)	(1.08)		
			Quantile Regressions for Hispanic Males				
BMI	-.0053	-.0049	-.0049	-0.0050	-.0034	6258	5
	(-3.61)	(-2.95)	(-3.52)	(-4.63)	(-2.13)		
BMI with lagged weight	-.0040	-.0055	-.0060	-.0071	-.0067	2819	6
	(-3.03)	(-3.74)	(-3.38)	(-4.90)	(-3.93)		

Notes:
) Data: NLSY79 males
2) For BMI and lagged 7 BMI, coefficients and t statistics are listed.
3) Other regressors include: number of children ever born, age of youngest child, general intelligence, highest grade completed, mother's highest grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, county unemployment rate, current school enrollment, part time job, white collar job, and region of residence.

Table 5.A1
Summary Statistics for Females NLSY 1981-2006 with censored regressors

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Adjusted wage CPI 2006	41002	14.52	22.38	1	852.08
Log wage	41002	2.46	0.61	0	6.75
Body mass index	41365	25.84	6.11	12.28	83.94
Weight in pounds corrected	41365	152.13	37.12	77.14	548.99
Height in inches corrected	41365	64.32	2.40	51.82	74.44
7-year lag BMI	28335	24.07	5.05	14.67	56.83
7-year lag weight	28335	141.42	31.02	81.33	319.51957
General intelligence	41365	44.21	27.13	1	99
Highest grade completed	41365	13.27	2.25	0	20
Mother's highest grade completed	41365	11.09	3.05	0	20
Father's highest grade completed	41365	11.05	3.83	0	20
Year	41365	8.83	4.42	2	16
Unemployment rate	41365	67.60	29.64	10	237
Number of years at current job	41365	4.08	4.80	0.02	37.74
Years of actual work experience	41365	7.61	7.37	0	30.78
Age	41365	30.99	7.41	16	49
Attending school	41365	0.18	0.38	0	1
Number of kids	41365	1.21	1.28	0	10
Age of youngest child	41365	2.09	6.75	-4	32
Hispanic	41365	0.16	0.37	0	1
Black	41365	0.24	0.42	0	1
Underweight	41365	0.03	0.18	0	1
Normal weight	41365	0.52	0.50	0	1
Overweight	41365	0.24	0.43	0	1

Table 5.A1
Summary Statistics for Females NLSY 1981-2006 with censored regressors (continued)

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Obese	41365	0.21	0.40	0	1
underw_7	28335	0.05	0.21	0	1
normalw_7	28335	0.59	0.49	0	1
overw_7	28335	0.21	0.41	0	1
obese_7	28335	0.15	0.36	0	1
Work more than 20 hours per week	41365	0.87	0.34	0	1
Work less than 20 hours per week	41365	0.13	0.34	0	1
White collar job	40082	0.64	0.48	0	1
Blue collar job	40082	0.36	0.48	0	1
Married spouse present	41365	0.50	0.50	0	1
Never married	41365	0.32	0.47	0	1
Married spouse not present	41365	0.19	0.39	0	1
Enrolled in school	41365	0.18	0.38	0	1
Not Enrolled in school	41365	0.82	0.38	0	1
Northeast region	41365	0.44	0.50	0	1
North Central region	41365	0.22	0.42	0	1
South region	41365	0.11	0.32	0	1
West region	41365	0.10	0.30	0	1
Have no kids	41365	0.01	0.11	0	1
Have kids between 0 and 5 years	41365	0.17	0.38	0	1
Have kids between 6 and 10	41365	0.25	0.43	0	1
Have kids between 11 and 18	41365	0.40	0.49	0	1
Have adult kids > 18	41365	0.19	0.39	0	1

Table 5.A2
Summary Statistics for Males NLSY 1981-2006 with Censored Regressors

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Adjusted wage CPI 2006	43962	18.44	27.08	1	852.08
Log wage	43962	2.68	0.62	0	6.75
Body mass index	44293	26.32	4.71	12.33	62.23
Weight in pounds corrected	44293	182.54	35.88	87.17	468.91
Height in inches corrected	44293	69.77	2.60	60.55	79.74
7-year lag BMI	30656	24.70	5.48	14.67	60.86
7-year lag weight	30656	145.21	33.49	81.33	351.91
General intelligence	44293	43.78	29.71	1	99
Highest grade completed	44293	12.97	2.42	3	20
Mother's highest grade completed	44293	11.11	3.17	0	20
Father's highest grade completed	44293	11.10	3.93	0	20
year	44293	8.69	4.38	2	16
Unemployment rate	44293	67.83	29.59	10	237
Number of years at current job	44293	4.35	5.03	0.02	32.02
Years of actual work experience	44293	7.62	7.77	0	30.64
Age	44293	30.71	7.30	16	49
Attending school	44293	0.14	0.35	0	1
Number of kids	44293	1.07	1.27	0	10
Age of youngest child	44293	-0.22	5.53	-4	29
Hispanic	44293	0.16	0.37	0	1
Black	44293	0.23	0.42	0	1
Underweight	44293	0.01	0.10	0	1
Normal weight	44293	0.44	0.50	0	1
Overweight	44293	0.37	0.48	0	1

Table 5.A2
Summary Statistics for Males NLSY 1981-2006 with Censored Regressors (continued)

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Obese	44293	0.18	0.39	0	1
underw_7	30656	0.02	0.13	0	1
normalw_7	30656	0.54	0.50	0	1
overw_7	30656	0.32	0.47	0	1
obese_7	30656	0.12	0.33	0	1
Work more than 20 hours per week	44293	0.94	0.24	0	1
Work less than 20 hours per week	44293	0.06	0.24	0	1
White collar job	42982	0.38	0.48	0	1
Blue collar job	42982	0.62	0.48	0	1
Married spouse present	44293	0.48	0.50	0	1
Never married	44293	0.40	0.49	0	1
Married spouse not present	44293	0.13	0.33	0	1
Enrolled in school	44293	0.14	0.35	0	1
Not Enrolled in school	44293	0.86	0.35	0	1
Northeast region	44293	0.59	0.49	0	1
North Central region	44293	0.20	0.40	0	1
South region	44293	0.07	0.25	0	1
West region	44293	0.05	0.21	0	1
Have no kids	44293	0.00	0.07	0	1
Have kids between 0 and 5 years	44293	0.18	0.38	0	1
Have kids between 6 and 10	44293	0.26	0.44	0	1
Have kids between 11 and 18	44293	0.36	0.48	0	1
Have adult kids > 18	44293	0.20	0.40	0	1

Table 5.A3

Summary Statistics for Females NLSY 1981-2006

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Adjusted wage CPI 2006	55112	14.09	22.02	1	852.08
Log wage	55112	2.42	0.63	0	6.75
Body mass index	55113	26.01	6.19	12.28	83.94
Weight in pounds corrected	55113	152.74	37.61	76.79	548.99
Height in inches corrected	55113	64.23	2.39	51.82	74.44
7-year lag BMI	37924	24.91	5.57	12.48	60.86
7-year lag weight	37924	146.05	34.10	76.79	354.15
General intelligence	55113	40.50	27.77	0	99
Highest grade completed	55113	13.09	2.33	0	20
Mother's highest grade completed	55113	10.40	3.84	0	20
Father's highest grade completed	55113	9.57	5.19	0	20
year	55113	9.04	4.53	2	16
Unemployment rate	55113	65.92	30.20	0	237
Number of years at current job	55113	3.85	4.69	0	37.74
Years of actual work experience	55113	7.32	7.42	0	31.56
Age	55113	31.21	7.66	16	49
Attending school	55113	0.17	0.38	0	1
Number of kids	55113	1.28	1.31	0	11
Age of youngest child	55113	4.06	5.49	0	32
Hispanic	55113	0.17	0.38	0	1
Black	55113	0.27	0.44	0	1
Underweight	55113	0.04	0.18	0	1
Normal weight	55113	0.50	0.50	0	1

Table 5.A3
Summary Statistics for Females NLSY 1981-2006 (continued)

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Overweight	55113	0.25	0.43	0	1
Obese	55113	0.22	0.41	0	1
underw_7	37924	0.05	0.21	0	1
normalw_7	37924	0.57	0.50	0	1
overw_7	37924	0.23	0.42	0	1
obese_7	37924	0.16	0.37	0	1
Work more than 20 hours per week	55113	0.85	0.36	0	1
Work less than 20 hours per week	55113	0.13	0.33	0	1
White collar job	55113	0.59	0.49	0	1
Blue collar job	55113	0.37	0.48	0	1
Married spouse present	55113	0.32	0.47	0	1
Never married	55113	0.48	0.50	0	1
Married spouse not present	55113	0.19	0.40	0	1
Enrolled in school	55113	0.17	0.38	0	1
Not Enrolled in school	55113	0.83	0.38	0	1
Northeast region	55113	0.17	0.37	0	1
North Central region	55113	0.23	0.42	0	1
South region	55113	0.41	0.49	0	1
West region	55113	0.19	0.39	0	1

Table 5.A4
Summary Statistics for Males NLSY 1981-2006

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Adjusted wage CPI 2006	60599	17.86	26.51	1	852.08
Log wage	60599	2.64	0.63	0	6.75
Body mass index	60602	26.31	4.80	10.44	66.72
Weight in pounds corrected	60602	181.98	36.34	81.94	468.91
Height in inches corrected	60602	69.68	2.63	60.53	79.78
7-year lag BMI	41354	25.23	4.41	12.83	61.07
7-year lag weight	41354	174.49	33.70	82.23	420.87
General intelligence	60602	38.94	30.30	0	99
Highest grade completed	60602	12.72	2.48	0	20
Mother's highest grade completed	60602	10.18	4.18	0	20
Father's highest grade completed	60602	9.53	5.30	0	20
year	60602	8.95	4.50	2	16
Unemployment rate	60602	66.38	30.40	0	237
Number of years at current job	60602	4.15	4.98	0	32.02
Years of actual work experience	60602	7.33	7.86	0	31.14
Age	60602	31.00	7.59	16	50
Attending school	60602	0.14	0.34	0	1
Number of kids	60602	1.12	1.31	0	10
Age of youngest child	60602	2.22	4.24	0	32
Hispanic	60602	0.18	0.38	0	1
Black	60602	0.26	0.44	0	1
Underweight	60602	0.01	0.10	0	1
Normal weight	60602	0.44	0.50	0	1
Overweight	60602	0.36	0.48	0	1

Table 5.A4
Summary Statistics for Males NLSY 1981-2006 (continued)

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Obese	60602	0.18	0.39	0	1
underw_7	41354	0.02	0.13	0	1
normalw_7	41354	0.54	0.50	0	1
overw_7	41354	0.32	0.47	0	1
obese_7	41354	0.13	0.33	0	1
Work more than 20 hours per week	60602	0.91	0.29	0	1
Work less than 20 hours per week	60602	0.06	0.24	0	1
White collar job	60602	0.34	0.47	0	1
Blue collar job	60602	0.62	0.48	0	1
Married spouse present	60602	0.40	0.49	0	1
Never married	60602	0.46	0.50	0	1
Married spouse not present	60602	0.14	0.34	0	1
Enrolled in school	60602	0.14	0.34	0	1
Not Enrolled in school	60602	0.86	0.34	0	1
Northeast region	60602	0.17	0.38	0	1
North Central region	60602	0.24	0.43	0	1
South region	60602	0.38	0.49	0	1
West region	60602	0.20	0.40	0	1

Table 5.A5

Coeficient and t-Statistic from Log Wage Regressions for Females 1981-2006 (Cawley 2004 replication)

	White Female			Black Female			Hispanic Female		
	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects
Column Number	1	2	3	4	5	6	7	8	9
BMI	-.0085	-.0095	-0.0052	-0.0031	-.0051	0.0028	-0.0069	-.0090	-0.0024
	(-6.83)	(-5.63)	(-3.12)	(-2.62)	(-3.22)	(1.7)	(-3.73)	(-3.94)	(-0.90)
Weight in pounds	-.0014	-0.0016	-0.0009	-.0005	-.0008	0.0002	-.0012	-.0016	-0.0006
	(-6.78)	(-5.64)	(-3.49)	(-2.56)	(-3.19)	(1.05)	(-3.71)	(-3.94)	(-1.37)
Underweight	0.0052	.0692	-0.0287	-.0888	-.0432	-0.1221	-.0417	-.0517	0.004
	(0.25)	(2.58)	(-1.3)	(-2.44)	(-1.22)	(-3.12)	(-1.19)	(-1.10)	(0.1)
Overweight	-.0400	-.0694	-0.0208	-.0030	-.0109	0.0328	-.0220	-.0384	0.0206
	(-2.97)	(-3.94)	(-1.54)	(-0.21)	(-0.64)	(2.24)	(-1.19)	(-1.54)	(0.97)
Obese	-.1285	-.1027	-0.085	-.0512	-.0755	0.0305	-.0860	-.1196	-0.0069
	(-6.61)	(-4.17)	(-3.97)	(-2.66)	(-3.26)	(1.46)	(-3.03)	(-3.58)	(-0.22)
Number of observations	30979	20665	30981	14741	10656	14743	9392	6602	9394

Notes:

1) Data: NLSY79 females

2) One of three measures of weight is used: BMI, weight in pounds (controlling for height in inches) or indicator variables for clinical weight classification: underweight, overweight, and obese (where healthy weight is the excluded categories).

3) For BMI and weight in pounds, coefficients and t-statistics are listed. For indicators of clinical weight classification, the percent change in log wages associated with a change in the indicator variables from 0 to 1 and t-statistics are listed.

4) Other regressors include: number of children ever born, age of youngest child, general intelligence, highest grade completed, mother's highest grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, county unemployment rate, current school enrollment, part time job, white collar job, and region of residence.

Table 5.A6

Coeficient and t-Statistic from Log Wage Regressions for Males 1981-2006 (Cawley 2004 replication)

	White Male			Black Male			Hispanic Male		
	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects
Column Number	1	2	3	4	5	6	7	8	9
BMI	-.0017	-.0037	-.0013	.0053	.0062	.0048	-.0053	-.0064	.0035
	(-1.06)	(-1.83)	(-0.63)	(2.78)	(2.51)	(1.97)	(-2.40)	(-2.33)	(1.26)
Weight in pounds	-.0002	-.0005	-.0004	.0007	.0008	.0005	-.0008	-.0010	.00009
	(-1.26)	(-2.03)	(-1.50)	(2.85)	(2.50)	(1.60)	(-2.56)	(-2.39)	(0.23)
Underweight	-.1609	-.0441	-.0496	-0.0460	-.1550	.0611	0.1288	.0507	.0824
	(-3.76)	(-0.91)	(-1.33)	(-1.36)	(-3.12)	(1.39)	(1.40)	(0.51)	(0.86)
Overweight	.0391	.0185	.0264	.0505	.0415	.0410	-0.004	.0042	.0265
	(2.92)	(1.18)	(2.29)	(3.28)	(2.25)	(2.55)	(-0.20)	(0.18)	(1.38)
Obese	-.0394	-.0838	-.0057	0.0665	.0432	.0540	-0.0321	-.0587	.0708
	(-1.98)	(-3.43)	(-0.30)	(2.83)	(1.38)	(2.24)	(-1.04)	(-1.56)	(2.24)
Number of observations	33733	22690	33735	16023	11111	16025	10843	7551	10845

Notes:

1) Data: NLSY79 males

2) One of three measures of weight is used: BMI, weight in pounds (controlling for height in inches) or indicator variables for clinical weight classification: underweight, overweight, and obese (where healthy weight is the excluded categories).

3) For BMI and weight in pounds, coefficients and t-tatistics are listed. For indicators of clinical weight classification, the percent change in log wages associated with a change in the indicator variables from 0 to 1 and t-statistics are listed.

4) Other regressors include: number of children ever born, age of youngest child, general intelligence, highest grade completed, mother's highest grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, county unemployment rate, current school enrollment, part time job, white collar job, and region of residence.

Table 5.A7

Coefficient and t-Statistic from Log Wage Quantile Regressions for Females 1981-2006

Percentile	20th	40th	50th	60th	80th	# Obs	Row Number
			Quantile Regressions for White Female				
BMI	-0.0071	-0.0073	-0.0077	-0.0078	-0.0083	30979	1
	(-10.54)	(-15.47)	(-15.32)	(-23.1)	(-16.57)		
BMI with lagged weight	-0.0077	-0.0088	-0.0092	-0.01	-0.0094	20665	2
	(-8.09)	(-11.5)	(-11.75)	(-10.62)	(-8.49)		
			Quantile Regressions for Black Females				
BMI	-0.0018	-0.0022	-0.0027	-0.0029	-0.0024	14741	3
	(-3.22)	(-4.45)	(-4.9)	(-5.02)	(-4.62)		
BMI with lagged weight	-0.0025	-0.0032	-0.0040	-0.0049	-0.0045	10656	4
	(-2.88)	(-3.45)	(-4.89)	(-7.04)	(-4.93)		
			Quantile Regressions for Hispanic Females				
BMI	-0.0038	-0.0038	-0.0050	-0.0055	-0.0067	9392	5
	(-3.88)	(-4.2)	(-6.00)	(-6.34)	(-6.03)		
BMI with lagged weight	-0.0054	-0.0061	-0.0069	-0.0080	-0.0102	6602	6
	(-3.84)	(-4.61)	(-4.90)	(-6.02)	(-7.13)		

Notes:

1) Data: NLSY79 females

2) For BMI and lagged 7 BMI, coefficients and t statistics are listed.

3) Other regressors include: number of children ever born, age of youngest child, general intelligence, highest grade completed, mother's highest grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, county unemployment rate, current school enrollment, part time job, white collar job, and region of residence.

Table 5.A8

Coefficients and t-Statistics from Log Wage Quantile Regressions for Males 1981-2006

Percentile	20th	40th	50th	60th	80th	# Obs	Row Number
			Quantile Regressions for White Males				
BMI	-.0013	-.0024	-.0022	-.0023	-.0005	33733	1
	(-1.62)	(-2.85)	(-2.69)	(-2.79)	(-0.48)		
BMI with lagged weight	-.0024	-.0031	-.0032	-.0031	-.0007	22690	2
	(-3.94)	(-3.47)	(-4.70)	(-4.39)	(-0.57)		
			Quantile Regressions for Black Males				
BMI	.0028	.0057	.0057	.0061	.0077	16023	3
	(2.11)	(6.95)	(8.76)	(6.65)	(6.55)		
BMI with lagged weight	.0039	.0056	.0054	.0053	.0082	11111	4
	(2.60)	(4.00)	(5.64)	(3.91)	(3.58)		
-			**Quantile Regressions for Hispanic Males**			-	-
BMI	-.0040	-.0037	-.0033	-.0035	-.0037	10843	5
	(-3.55)	(-3.84)	(-2.98)	(-3.42)	(-2.28)		
BMI with lagged weight	-.0045	-.0038	-.0056	-.0055	-.0052	7551	6
	(-2.95)	(-3.69)	(-5.79)	(-5.38)	(-.0052)		

Notes:

1) Data: NLSY79 males

2) For BMI and lagged 7 BMI, coefficients and t statistics are listed.

3) Other regressors include: number of children ever born, age of youngest child, general intelligence, highest grade completed, mother's highest grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, county unemployment rate, current school enrollment, part time job, white collar job, and region of residence.

Table 5.A9

Quantile Regression of the Wage Model for White Females and Black Males Using Censored Regressors Data NLSY 1979-2006 (More regressions are available on demand)

gender=female, race=white

	# of obs	24099
0.2	Pseudo R2 =	0.1675
0.4	Pseudo R2 =	0.225
0.5	Pseudo R2 =	0.2387
0.6	Pseudo R2 =	0.2472
0.8	Pseudo R2 =	0.249

logwage	Coefficient	Robust Std. Err.	t-statistic	P>|t|	[95% Conf.Interval]	[95% Conf.Interval]
q20 |						
Body mass index	-0.0074944	0.000703	-10.66	0	-0.00887	-0.00612
Age	-0.0045324	0.001915	-2.37	0.018	-0.00828	-0.00078
Time	0.0198733	0.002686	7.4	0	0.014609	0.025138
Married spouse present	0.0271217	0.011468	2.36	0.018	0.004644	0.0496
Married spouse not present	0.0489174	0.012041	4.06	0	0.025316	0.072519
Highest grade Completed	0.0370149	0.001985	18.65	0	0.033125	0.040905
Attending school	-0.0908674	0.009058	-10.03	0	-0.10862	-0.07311
Highest grade Compl by mother	0.0036252	0.001642	2.21	0.027	0.000407	0.006844

Highest grade Compl by father	-0.0008303	0.001387	-0.6	0.55	-0.00355	0.001889
General intelligence	0.0026227	0.000186	14.11	0	0.002258	0.002987
Work less than 20 hours per week	-0.1674599	0.014792	-11.32	0	-0.19645	-0.13847
Tenure	0.0275374	0.001192	23.11	0	0.025202	0.029873
Years of work experience	0.0059632	0.000677	8.81	0	0.004637	0.00729
County unemployment rate	-0.0015908	0.000155	-10.28	0	-0.00189	-0.00129
White collar	0.1754688	0.009104	19.27	0	0.157625	0.193313
Number of Kids	-0.0508978	0.005123	-9.93	0	-0.06094	-0.04086
Have kids between 0 and 5 years	0.0074543	0.009795	0.76	0.447	-0.01174	0.026653
Have kids between 6 and 10	-0.0275035	0.011647	-2.36	0.018	-0.05033	-0.00468
Have kids between 11 and 18	-0.0596405	0.018793	-3.17	0.002	-0.09648	-0.0228
Have adult kids > 18	-0.016204	0.022389	-0.72	0.469	-0.06009	0.027681
Northeast region	0.0989923	0.012942	7.65	0	0.073625	0.12436
North Central Region	-0.0094614	0.009182	-1.03	0.303	-0.02746	0.008537
West region	0.0842877	0.009467	8.9	0	0.065731	0.102844
Constant	1.56388	0.046044	33.96	0	1.47363	1.65413
-------------+-	------------	-----------	---------	---------	------------	----------
q40 |						
Body mass index	-0.0080468	0.000555	-14.5	0	-0.00913	-0.00696
Age	0.000193	0.000963	0.2	0.841	-0.00169	0.00208
Time	0.01833	0.001908	9.61	0	0.01459	0.02207
Married spouse present	0.0427519	0.007991	5.35	0	0.027089	0.058415
Married spouse not present	0.0316491	0.009659	3.28	0.001	0.012718	0.050581
Highest grade Completed	0.050206	0.002077	24.17	0	0.046135	0.054277

Attending school	-0.086891	0.007764	-11.19	0	-0.10211	-0.07167
Highest grade Compl by mother	0.0005311	0.001414	0.38	0.707	-0.00224	0.003303
Highest grade Compl by father	0.0010072	0.001017	0.99	0.322	-0.00099	0.003001
General intelligence	0.0028101	0.000201	13.96	0	0.002416	0.003205
Work less than 20 hours per week	-0.1419928	0.014907	-9.53	0	-0.17121	-0.11277
Tenure	0.0268834	0.000969	27.74	0	0.024984	0.028783
Years of work experience	0.0046602	0.000639	7.3	0	0.003408	0.005912
County unemployment rate	-0.001556	8.94E-05	-17.41	0	-0.00173	-0.00138
White collar	0.1338551	0.007694	17.4	0	0.118775	0.148935
Number of Kids	-0.0383694	0.004195	-9.15	0	-0.04659	-0.03015
Have kids between 0 and 5 years	-0.0072133	0.012762	-0.57	0.572	-0.03223	0.017801
Have kids between 6 and 10	-0.0685575	0.015615	-4.39	0	-0.09916	-0.03795
Have kids between 11 and 18	-0.0924433	0.012541	-7.37	0	-0.11702	-0.06786
Have adult kids > 18	-0.1034676	0.029341	-3.53	0	-0.16098	-0.04596
Northeast region	0.1296807	0.012714	10.2	0	0.104761	0.1546
North Central Region	0.0057392	0.006979	0.82	0.411	-0.00794	0.019419
West region	0.1299466	0.010375	12.52	0	0.109611	0.150283
Constant	1.498965	0.031294	47.9	0	1.437627	1.560302
-------------+-	------------	-----------	---------	---------	------------	----------
q50 |						
Body mass index	-0.0080678	0.000566	-14.25	0	-0.00918	-0.00696
Age	0.0011773	0.000971	1.21	0.225	-0.00073	0.00308
Time	0.0192984	0.001265	15.26	0	0.016819	0.021778
Married spouse present	0.041841	0.006844	6.11	0	0.028426	0.055256

Married spouse not present	0.0382152	0.01003	3.81	0	0.018555	0.057875
Highest grade Completed	0.0557807	0.002177	25.63	0	0.051515	0.060047
Attending school	-0.0879321	0.007304	-12.04	0	-0.10225	-0.07362
Highest grade Compl by mother	-0.0006681	0.001263	-0.53	0.597	-0.00314	0.001808
Highest grade Compl by father	0.0024003	0.000878	2.73	0.006	0.00068	0.004121
General intelligence	0.002893	0.000169	17.15	0	0.002562	0.003224
Work less than 20 hours per week	-0.1225066	0.010426	-11.75	0	-0.14294	-0.10207
Tenure	0.0264832	0.000929	28.49	0	0.024661	0.028305
Years of work experience	0.0046052	0.000559	8.24	0	0.00351	0.0057
County unemployment rate	-0.0016155	9.46E-05	-17.07	0	-0.0018	-0.00143
White collar	0.1183313	0.0068	17.4	0	0.105004	0.131659
Number of Kids	-0.0397932	0.003669	-10.85	0	-0.04698	-0.0326
Have kids between 0 and 5 years	-0.0015133	0.01048	-0.14	0.885	-0.02205	0.019028
Have kids between 6 and 10	-0.0501769	0.010357	-4.84	0	-0.07048	-0.02988
Have kids between 11 and 18	-0.0903415	0.013489	-6.7	0	-0.11678	-0.0639
Have adult kids > 18	-0.1235376	0.040945	-3.02	0.003	-0.20379	-0.04328
Northeast region	0.1501642	0.010977	13.68	0	0.128649	0.17168
North Central Region	0.0121749	0.006945	1.75	0.08	-0.00144	0.025787
West region	0.1412348	0.009149	15.44	0	0.123303	0.159166
Constant	1.479498	0.018943	78.1	0	1.442368	1.516627
-------------+-	------------	-----------	---------	---------	------------	----------
q60 |						
Body mass index	-0.0080925	0.000562	-14.4	0	-0.00919	-0.00699
Age	0.0006923	0.001463	0.47	0.636	-0.00217	0.003559

Time	0.0232193	0.001976	11.75	0	0.019346	0.027092
Married spouse present	0.0344796	0.01009	3.42	0.001	0.014704	0.054256
Married spouse not present	0.0275665	0.011396	2.42	0.016	0.00523	0.049903
Highest grade Completed	0.0614699	0.002301	26.72	0	0.05696	0.06598
Attending school	-0.0954229	0.008606	-11.09	0	-0.11229	-0.07855
Highest grade Compl by mother	-0.0022649	0.001278	-1.77	0.076	-0.00477	0.00024
Highest grade Compl by father	0.0037563	0.000877	4.28	0	0.002038	0.005475
General intelligence	0.0030131	0.000195	15.43	0	0.00263	0.003396
Work less than 20 hours per week	-0.0984668	0.011777	-8.36	0	-0.12155	-0.07538
Tenure	0.0259072	0.000899	28.81	0	0.024144	0.02767
Years of work experience	0.0045531	0.000796	5.72	0	0.002994	0.006112
County unemployment rate	-0.0015396	0.000098	-15.7	0	-0.00173	-0.00135
White collar	0.1133684	0.007336	15.45	0	0.09899	0.127747
Number of Kids	-0.0348392	0.003571	-9.76	0	-0.04184	-0.02784
Have kids between 0 and 5 years	0.0088372	0.01091	0.81	0.418	-0.01255	0.030221
Have kids between 6 and 10	-0.0520643	0.014007	-3.72	0	-0.07952	-0.02461
Have kids between 11 and 18	-0.099835	0.014892	-6.7	0	-0.12902	-0.07065
Have adult kids > 18	-0.1294398	0.032491	-3.98	0	-0.19312	-0.06576
Northeast region	0.1508764	0.011172	13.51	0	0.128979	0.172774
North Central Region	0.0138282	0.008935	1.55	0.122	-0.00369	0.031342
West region	0.1447207	0.011245	12.87	0	0.12268	0.166762
Constant	1.479437	0.025617	57.75	0	1.429226	1.529647
-------------+-	------------	-----------	---------	---------	------------	----------
q80 |						

Body mass index	-0.0084175	0.000717	-11.74	0	-0.00982	-0.00701
Age	0.0041601	0.001642	2.53	0.011	0.000942	0.007378
Time	0.0277949	0.002784	9.98	0	0.022338	0.033252
Married spouse present	0.0377558	0.011238	3.36	0.001	0.01573	0.059782
Married spouse not present	0.0232711	0.011759	1.98	0.048	0.000222	0.04632
Highest grade Completed	0.0720445	0.002828	25.48	0	0.066503	0.077587
Attending school	-0.1097837	0.007386	-14.86	0	-0.12426	-0.09531
Highest grade Compl by mother	-0.0042308	0.001527	-2.77	0.006	-0.00722	-0.00124
Highest grade Compl by father	0.0065822	0.001367	4.82	0	0.003904	0.009261
General intelligence	0.0028155	0.000214	13.17	0	0.002397	0.003234
Work less than 20 hours per week	-0.0145201	0.014723	-0.99	0.324	-0.04338	0.014337
Tenure	0.022165	0.000862	25.72	0	0.020476	0.023854
Years of work experience	0.0035705	0.000711	5.02	0	0.002177	0.004965
County unemployment rate	-0.0014577	0.000123	-11.87	0	-0.0017	-0.00122
White collar	0.0909195	0.008293	10.96	0	0.074665	0.107174
Number of Kids	-0.0391003	0.004756	-8.22	0	-0.04842	-0.02978
Have kids between 0 and 5 years	0.0325648	0.011224	2.9	0.004	0.010566	0.054564
Have kids between 6 and 10	-0.0528337	0.015382	-3.43	0.001	-0.08298	-0.02268
Have kids between 11 and 18	-0.1288208	0.018976	-6.79	0	-0.16601	-0.09163
Have adult kids > 18	-0.1789855	0.040065	-4.47	0	-0.25751	-0.10046
Northeast region	0.1333566	0.006912	19.29	0	0.119809	0.146905
North Central Region	0.0076831	0.010418	0.74	0.461	-0.01274	0.028104
West region	0.1521979	0.012359	12.31	0	0.127974	0.176422
Constant	1.46852	0.045969	31.95	0	1.378418	1.558621

---------------- ------------ ----------- --------- --------- ------------ ----------

gender=male, race=Black

	# of obs	9681
0.2	Pseudo R2 =	0.1717
0.4	Pseudo R2 =	0.2198
0.5	Pseudo R2 =	0.2321
0.6	Pseudo R2 =	0.238
0.8	Pseudo R2 =	0.2293

logwage	**Coefficient**	**Robust Std. Err.**	**t-statistic**	**P>|t|**	**[95% Conf.Interval]**	**[95% Conf.Interval]**
q20 |						
Body mass index	0.0018504	0.001434	1.29	0.197	-0.00096	0.004661
Age	-0.011218	0.002187	-5.13	0	-0.0155	-0.00693
Time	0.0087983	0.002839	3.1	0.002	0.003233	0.014364
Married spouse present	0.1750973	0.021813	8.03	0	0.13234	0.217855
Married spouse not present	0.0730869	0.017853	4.09	0	0.038092	0.108082
Highest grade Completed	0.0476692	0.003982	11.97	0	0.039864	0.055475
Attending school	-0.1466046	0.019191	-7.64	0	-0.18422	-0.10899

Highest grade Compl by mother	0.0003911	0.002885	0.14	0.892	-0.00526	0.006046
Highest grade Compl by father	0.0001626	0.001703	0.1	0.924	-0.00318	0.003501
General intelligence	0.0036896	0.000268	13.76	0	0.003164	0.004215
Work less than 20 hours per week	-0.1873357	0.024361	-7.69	0	-0.23509	-0.13958
Tenure	0.0293482	0.002044	14.36	0	0.025342	0.033355
Years of work experience	0.005436	0.001151	4.72	0	0.00318	0.007692
County unemployment rate	-0.0018137	0.00024	-7.55	0	-0.00228	-0.00134
White collar	0.0578991	0.014267	4.06	0	0.029933	0.085865
Number of Kids	-0.0044545	0.004359	-1.02	0.307	-0.013	0.004091
Have kids between 0 and 5 years	0.0275909	0.013754	2.01	0.045	0.00063	0.054552
Have kids between 6 and 10	0.0464278	0.034507	1.35	0.179	-0.02121	0.114069
Have kids between 11 and 18	0.0716376	0.027308	2.62	0.009	0.018109	0.125167
Have adult kids > 18	0.1212623	0.050104	2.42	0.016	0.023048	0.219477
Northeast region	0.128782	0.0153	8.42	0	0.098791	0.158774
North Central Region	0.0135606	0.015008	0.9	0.366	-0.01586	0.04298
West region	0.15466	0.019424	7.96	0	0.116585	0.192735
Constant	1.603015	0.100752	15.91	0	1.405519	1.80051
-------------+-	------------	-----------	---------	---------	------------	----------
q40 |						
Body mass index	0.0030596	0.000915	3.34	0.001	0.001266	0.004853
Age	-0.0050581	0.00182	-2.78	0.005	-0.00863	-0.00149
Time	0.0047905	0.002996	1.6	0.11	-0.00108	0.010663
Married spouse present	0.1609979	0.017396	9.25	0	0.126897	0.195099
Married spouse not present	0.0920694	0.012828	7.18	0	0.066923	0.117216

Highest grade Completed	0.0486565	0.003091	15.74	0	0.042597	0.054716
Attending school	-0.1412353	0.016717	-8.45	0	-0.174	-0.10847
Highest grade Compl by mother	0.0033647	0.002192	1.54	0.125	-0.00093	0.007661
Highest grade Compl by father	0.0033434	0.0014	2.39	0.017	0.000599	0.006088
General intelligence	0.0043361	0.000311	13.93	0	0.003726	0.004946
Work less than 20 hours per week	-0.1483035	0.026481	-5.6	0	-0.20021	-0.0964
Tenure	0.0291122	0.001441	20.21	0	0.026288	0.031936
Years of work experience	0.0048553	0.00061	7.96	0	0.003659	0.006051
County unemployment rate	-0.0016184	0.00028	-5.77	0	-0.00217	-0.00107
White collar	0.0693602	0.011893	5.83	0	0.046048	0.092672
Number of Kids	-0.0055859	0.003828	-1.46	0.145	-0.01309	0.001917
Have kids between 0 and 5 years	0.0404174	0.018668	2.17	0.03	0.003824	0.077011
Have kids between 6 and 10	0.0675658	0.031175	2.17	0.03	0.006456	0.128676
Have kids between 11 and 18	0.0506818	0.029082	1.74	0.081	-0.00632	0.107688
Have adult kids > 18	0.0382818	0.058872	0.65	0.516	-0.07712	0.153684
Northeast region	0.1463759	0.012713	11.51	0	0.121456	0.171296
North Central Region	0.0033827	0.017529	0.19	0.847	-0.03098	0.037743
West region	0.1529321	0.021747	7.03	0	0.110304	0.19556
Constant	1.517676	0.067651	22.43	0	1.385066	1.650286
-------------+-	------------	-----------	---------	---------	------------	----------
q50 |						
Body mass index	0.0027167	0.000819	3.32	0.001	0.001111	0.004323
Age	-0.0043244	0.001888	-2.29	0.022	-0.00803	-0.00062
Time	0.0051337	0.002852	1.8	0.072	-0.00046	0.010725

Married spouse present	0.1617343	0.020203	8.01	0	0.122133	0.201336
Married spouse not present	0.1033522	0.013032	7.93	0	0.077808	0.128897
Highest grade Completed	0.0540005	0.002834	19.05	0	0.048445	0.059556
Attending school	-0.1369964	0.014482	-9.46	0	-0.16538	-0.10861
Highest grade Compl by mother	0.0054755	0.001654	3.31	0.001	0.002233	0.008718
Highest grade Compl by father	0.0030557	0.001342	2.28	0.023	0.000425	0.005687
General intelligence	0.0045228	0.000274	16.5	0	0.003985	0.00506
Work less than 20 hours per week	-0.1175277	0.019575	-6	0	-0.1559	-0.07916
Tenure	0.0270406	0.001177	22.97	0	0.024733	0.029348
Years of work experience	0.0049823	0.000672	7.41	0	0.003664	0.0063
County unemployment rate	-0.00171	0.000196	-8.71	0	-0.0021	-0.00133
White collar	0.0691794	0.011273	6.14	0	0.047081	0.091278
Number of Kids	-0.0059676	0.003746	-1.59	0.111	-0.01331	0.001375
Have kids between 0 and 5 years	0.0493383	0.015108	3.27	0.001	0.019723	0.078954
Have kids between 6 and 10	0.0580836	0.026762	2.17	0.03	0.005625	0.110542
Have kids between 11 and 18	0.0721518	0.021342	3.38	0.001	0.030317	0.113987
Have adult kids > 18	-0.0078966	0.105616	-0.07	0.94	-0.21493	0.199132
Northeast region	0.1469154	0.013269	11.07	0	0.120906	0.172924
North Central Region	-0.0044745	0.014725	-0.3	0.761	-0.03334	0.02439
West region	0.1559178	0.015855	9.83	0	0.124839	0.186997
Constant	1.505776	0.056385	26.71	0	1.395251	1.616302
-------------+-	------------	-----------	---------	---------	------------	----------
q60 |						
Body mass index	0.0027675	0.000821	3.37	0.001	0.001158	0.004377

Age	-0.0024224	0.002509	-0.97	0.334	-0.00734	0.002496
Time	0.0039759	0.003233	1.23	0.219	-0.00236	0.010314
Married spouse present	0.1582792	0.018608	8.51	0	0.121803	0.194756
Married spouse not present	0.1207402	0.014457	8.35	0	0.092402	0.149079
Highest grade Completed	0.0542145	0.00283	19.16	0	0.048668	0.059761
Attending school	-0.1424242	0.01439	-9.9	0	-0.17063	-0.11422
Highest grade Compl by mother	0.0076698	0.00211	3.63	0	0.003534	0.011806
Highest grade Compl by father	0.0012184	0.001521	0.8	0.423	-0.00176	0.004199
General intelligence	0.0047616	0.000335	14.23	0	0.004106	0.005418
Work less than 20 hours per week	-0.1104889	0.020379	-5.42	0	-0.15044	-0.07054
Tenure	0.0267416	0.001628	16.42	0	0.02355	0.029934
Years of work experience	0.0049106	0.000795	6.18	0	0.003353	0.006469
County unemployment rate	-0.0016652	0.000234	-7.11	0	-0.00212	-0.00121
White collar	0.0738997	0.010452	7.07	0	0.053412	0.094387
Number of Kids	-0.0075354	0.003507	-2.15	0.032	-0.01441	-0.00066
Have kids between 0 and 5 years	0.0506962	0.016047	3.16	0.002	0.019242	0.082151
Have kids between 6 and 10	0.0904614	0.029163	3.1	0.002	0.033296	0.147627
Have kids between 11 and 18	0.0764266	0.025255	3.03	0.002	0.026921	0.125932
Have adult kids > 18	-0.0070862	0.11854	-0.06	0.952	-0.23945	0.225277
Northeast region	0.1520286	0.017703	8.59	0	0.117328	0.186729
North Central Region	0.0108699	0.017756	0.61	0.54	-0.02394	0.045675
West region	0.183934	0.015056	12.22	0	0.154421	0.213447
Constant	1.525783	0.053422	28.56	0	1.421064	1.630502
-------------+-	------------	-----------	---------	---------	------------	----------

q80 |						
Body mass index	0.0035291	0.001104	3.2	0.001	0.001366	0.005692
Age	0.0065861	0.002814	2.34	0.019	0.00107	0.012102
Time	-0.0040977	0.004171	-0.98	0.326	-0.01227	0.004079
Married spouse present	0.1289042	0.017233	7.48	0	0.095123	0.162685
Married spouse not present	0.1238494	0.017595	7.04	0	0.08936	0.158339
Highest grade Completed	0.0556004	0.004177	13.31	0	0.047412	0.063788
Attending school	-0.1646488	0.019105	-8.62	0	-0.2021	-0.1272
Highest grade Compl by mother	0.0074663	0.003216	2.32	0.02	0.001162	0.01377
Highest grade Compl by father	-0.0002675	0.00246	-0.11	0.913	-0.00509	0.004555
General intelligence	0.0048207	0.000471	10.23	0	0.003897	0.005744
Work less than 20 hours per week	0.0287114	0.042512	0.68	0.499	-0.05462	0.112043
Tenure	0.02178	0.00127	17.16	0	0.019291	0.024269
Years of work experience	0.0022702	0.000866	2.62	0.009	0.000573	0.003967
County unemployment rate	-0.0013651	0.00026	-5.24	0	-0.00188	-0.00085
White collar	0.0913833	0.012381	7.38	0	0.067114	0.115653
Number of Kids	-0.008825	0.005729	-1.54	0.123	-0.02006	0.002405
Have kids between 0 and 5 years	0.064266	0.023776	2.7	0.007	0.01766	0.110872
Have kids between 6 and 10	0.046624	0.019516	2.39	0.017	0.008369	0.08488
Have kids between 11 and 18	0.081913	0.031331	2.61	0.009	0.020498	0.143329
Have adult kids > 18	0.0143551	0.105632	0.14	0.892	-0.19271	0.221415
Northeast region	0.1867194	0.016543	11.29	0	0.154291	0.219148
North Central Region	0.0423801	0.017133	2.47	0.013	0.008795	0.075965
West region	0.2462972	0.023781	10.36	0	0.199681	0.292913

Constant	1.529775	0.05784	26.45	0	1.416397	1.643152
---------------	------------	-----------	---------	---------	------------	----------

Chapter 6

DO THE EFFECTS OF WEIGHT ON WAGES CHANGE OVER TIME?

In chapters four and five, I have analyzed the effects of obesity for persons who were part of the NLSY79 over the period 1981 through 2006. This chapter utilizes two data sets, the National Longitudinal Survey of Youth in 1979 and the National Longitudinal Survey of Youth in 1997. The objective is to determine whether the effects of obesity have changed over time. From each data set we are able to estimate the effects of weight on wages for a group of respondents who are the same ages (16 to 29) in two different time periods. The first time period reflects the 1980s while the second includes the early years of the twenty first century. My results provide some evidence that the effects of weight on wages decrease for white women over the two decades. We are unaware of any studies that have attempted to look at changes in the weight penalty using comparable data sets (NLSY79 and NLSY97) and the same age cohorts.

The outline of this chapter is as follows. Section 6.1 is an introduction to the chapter. Section 6.2 discusses the descriptive statistics and the important empirical results (by gender and ethnicity) regarding changes over time in the effects of weight on wages. Conclusions are included in section 6.3.

6.1 Introduction

Since the effect of obesity on wages is of specific interest, especially if obesity is a factor in hiring and/or promotions, it is very important to understand not only the level of the wage penalty, but also if and how the penalty has changed over time. This chapter utilizes two data sets to see if there have been changes in the effects of weight on wages. The first data set is the National Longitudinal Survey of Youth 1979 (NLSY79) for respondents aged between 16-29 (born between 1957 and 1964) and the second data sets is National Longitudinal Survey of Youth 1997 (NLSY97) also aged between 16-29 (born between 1980 and 1984). Both data sets are created and specified as described in chapter 3[1].

As in chapters 4 and 5 the correlation between weight and wages could in fact reflect both that body weight affects wages and/or that wages affect body weight. We again follow Cawley (2000, 2004) and Cawley et al. (2005), and replace the contemporaneous weight with its 2-years lagged value[2] in the OLS regression

2. Two years lagged is used instead of 7 because of the availability of the data

and then estimate fixed effects models to control for unobserved individual effects.

Also as in earlier chapters, we follow the procedure suggested by (Rigobon and Stoker (2009)) for handling missing data in the regressors, censored regressors. We are unaware of any other studies of the obesity-wage connection that have used this approach. Nor are we aware of any studies of the issue at hand that have taken cohorts of the same age at different time periods to check for changes in the effect of obesity on wages over time.

I find that, although negative correlations between weight and wages are observed for white women in both cohorts, white women in the NLSY97 appear to be subject to a smaller penalty associated with being overweight.

6.2 Empirical Results

6.2.1 Descriptive Statistics

Appendices Tables 6A1, 6A2, 6A3, and 6A4 provide summary statistics by gender. Table 6A1 and 6.A2 provide descriptive statistics for the samples of females and males respectively, of the NLSY79 cohort for the respondents age 16 to 29. Tables 6.A3 and 6.A4 present the descriptive statistics for the samples of females and males respectively of the NLSY97 cohort for the respondents age 16 to 29. We next compare the means for some key variables by gender. The mean age for the samples NLSY79 for both males and females is 24, and the mean age for the samples NLSY97 for both males and females is 22. We will examine means for females. The BMI increases from 24 to 26, and weight increases from 141 to 156 pounds over time. Means for mother's highest grade completed, father's highest grade completed, and general intelligence were for the NLSY79 sample (NLSY97 sample) respectively 11.18 (13.09), 11.19 (13.6) and 44 (52). The mean of height in all samples is 64 inches for both cohorts. The cohort NLSY97 is more obese and more educated.

As with the females, most of the means for males are greater in the NLSY97. The BMI increases from 24 to 26, and weight increases from 171 to 182 pounds. Means for mother's highest grade completed, father's highest grade completed and general intelligence were for the NLSY79 sample (NLSY97 sample) respectively 11.06 (13.09), 11.06 (13.08) and 42 (47). The exception is for the height. In all samples the mean height is 70 inches for both cohorts. Males of the NLSY97 cohort are also heavier and more educated.

6.2.2 Econometric Methods and Analysis

Following Rigobon and Stoker (2009), I estimated all the OLS and fixed effect regressions mentioned in chapter 3 with censored regressors (deleting all observations that had missing values for the explanatory variables instead of creating dummy variables to account for missing variables). Using two data sets NLSY79 and NLSY97 for respondents aged 16-29, I compare the results of the wage equations for the two cohorts. The first cohort includes people born between 1957 and 1964 (baby boomers) while the second includes people born during the years 1980 through 1984.

Tables 6.1, 6.2, 6.3, and 6.4 present the estimated OLS coefficients (along with accompanying t-statistics) and fixed effect coefficients respectively using the model for log hourly wages, W, for females and males by race and ethnicity:

$$LnW_{it} = X_{it}\gamma + \varepsilon_{it} \qquad (3.1)$$

where the subscript i is for the individual, t for time, X_{it} a vector of explanatory variables that include the variable BMI (or another variables measuring weight) that affect wages, and ε_{it} is the residual for observation i in time t (for more details see chapter 3 section 3.6.2). The regressors are the same as were used in Chapter 4 and 5 and are defined in section 3.3. The complete list of variables that are included in addition to the weight variables are shown below the tables.

For the NLSY79 cohort, the OLS results in columns 1, 4, and 7 of Table 6.1 indicate that, for each ethnic group of females, both BMI and weight in pounds have negative coefficients but are only statistically significant for whites and Blacks. Although the sample size used in this chapter is much smaller than the one used in chapter 4 (35,387 observations for females) because we used only respondents aged between 16 and 29 from the NLSY79 versus all ages in chapter 4, the coefficients of the lagged wages on weight are similar and statistically significant.

For the NLSY97 female cohort, the OLS results in columns 1, 4, and 7 of Table 6.3 indicate that, although the coefficients for BMI and weight are negative for white females, they are never significant at the 5 percent level of significance. The coefficients in Table 6.3 and the standard deviations of weight in pounds in Table 6.A3 imply that for two identical white women who differ in weight by two standard deviations (roughly 75 pounds), we would expect the lighter one to earn 3.7 percent higher wages than the heavier one. The effect for the cohort of NLSY79 is associated with 8.7. There is a difference which is roughly 5.5 percent between the two coefficients. The negative effect of weight on wages appears to have decreased over time when we look at the BMI and weight in pounds.

Columns 1, 4, and 7 of Table 6.2 show results for OLS models for males from the NLSY79. Results for BMI and contemporaneous weight indicate that their coefficients in the OLS models are not significant. The results for NLSY97 in columns 1, 4, and 7 of Table 6.4, indicate that all the coefficients for BMI and contemporaneous weight are not significant.

For respondents in the NLSY79 data, columns 1, 4, and 7 of Tables 6.1 and 6.2 also present the OLS coefficients on the dummy variables for clinical weight classifications (underweight, healthy weight, overweight, and obese where healthy weight is the reference group). Among white (Black) females, those who are overweight earn 6.5 (3.8) percent less than those having healthy weight, and those who are obese earn 11.7 (6.3) percent less than those having a healthy weight. Overweight Hispanic females earn no less than the persons classified as having healthy weight. However obese Hispanic females earn almost 8 percent less than their counterparts who have a healthy weight.

Columns 1, 4, and 7 of Tables 6.3 and 6.4 present the OLS coefficients on the dummy variables for clinical weight classifications for the respondents born between 1980 and 1984 (NLSY97). Most of the coefficients were not significant for either females or males. The exception is for underweight Black (Hispanic) males who are subject to a wage penalty of 10.7 (17.8) respectively when compared to a similarly situated male having a healthy weight. Hispanic females are subject to a wage penalty of 19.2 percent.

The weight variables and error term in equation 3.1 may be correlated if there is reverse causality or if an unobserved variable causes either higher weight and/or adverse labor market outcomes. One of the approaches Cawley uses to get at the endogeneity problem is by replacing contemporaneous weight with its 7-year lagged value in the OLS regressions. For members of the NLSY79 cohorts, columns 2, 5, and 8 in Tables 6.1 and 6.2 present OLS results using a measure of weight lagged two years. Due to the availability of data only a two year lag is possible. The results for BMI and weight in pounds are shown in rows 1 and 2. For white females the OLS estimates of the contemporaneous measures of the weight are similar to the results based on lagged measures with the coefficients indicating wage penalties associated with being heavier. The coefficients for Black females are negative in both OLS and lagged models but are no longer significant once the lag is introduced. The coefficients for Hispanic females although negative, are not significant in either the OLS contemporaneous weight models or OLS with lagged variable models.

For females, born between 1980 and 1984 and included in the NLSY97, Columns 2, 5, and 8 in Tables 6.3 present OLS results using a measure of weight lagged two years. For white females, the OLS estimates of lagged measures of the weight variable are similar with all coefficients of the lagged measures of the baby boomer generation (Table 6.1). If we compare row 1 column 1 of the tables 6.1 and

6.3 we see that the OLS coefficients are smaller for results based on the NLSY97. However when looking at the lagged models the coefficients in the two tables are almost the same.

One of the most compelling results is that, when using the lagged measure of weight and BMI with the data based on the NLSY97, the wage penalty associated with weight disappears for Black and Hispanic females.

Table 6.4 shows the lagged results for contemporaneous BMI and weight for males included in NLSY97. None of the coefficients is significantly different from zero. For lagged weight dummy variables a negative and significant effect of obesity on wages was found for males in the NLSY79 data but the effect disappears in the more recent cohort of males in the NLSY97.

Cawley estimates fixed effect models to control for unobservable individual effects. Following Cawley, I estimate fixed-effects models to eliminate time-invariant heterogeneity. We exploit the panel nature of the NLSY79 and NLSY97 data sets to eliminate individual-specific fixed effects, assuming that the influence of genes and non-genetic factors is constant over time. Columns 3, 6, and 9 of Tables 6.1 and 6.2 report estimates from fixed-effects regressions for females and males, respectively, who participated in the NLSY79. Tables 6.3 and 6.4 present fixed effects models for the cohort from NLSY97. For white females, results from the fixed effects models show that the coefficients on BMI and weight in pounds are negative and statistically significant for the young cohort which was included in the NLSY79. These effects disappear for young white women in the NLSY97. Another notable difference occurs with Hispanic females where the coefficients on the dummy variables 'overweight' and 'obese' are negative and significant in the NLSY79 data but not in the NLSY97 data.

Columns 3, 6, and 9 of Tables 6.3 and 6.4 report estimates from fixed-effects regressions for the NLSY97. As the Tables show none of the coefficients is statistically significant for males or females. As Cawley specifies, the effect of weight on wages is maybe due to unobserved heterogeneity.

6.3 Conclusions:

This chapter investigates the change of the effect of weight on wage in the U.S. over time. For the purpose, I compare the effects of weight on wages between two generations aged 16 to 29. The first generation, that I called baby boomers (born between 1957 and 1964), is taken from the National Longitudinal Survey of Youth 1979 (NLSY79). The second cohort is taken from the National Longitudinal Survey of Youth 1997 (NLSY97) where respondents were born between 1980 and 1989.

Ordinary least squares results indicate that heavier white females, Black females, and Hispanic females, born between 1957 and 1964, tend to earn less than their lighter counterparts. Models also are estimated using lagged body weight, in order to account for the possible influence of wages on contemporaneous weight. The results for the cohort based on NLSY79 data in chapter 4 are similar to the results for the young cohort based on NLSY79. For example, if we take two identical white female born between 1957 and 1964, aged between 16 and 29 and who differ in weight by two standard deviations (roughly 58 pounds), we would expect the lighter one to earn wages that are higher by 8.12 percent higher wages.

However the results for females born between 1980 and 1984 (NLSY97) are different from the results for females born between 1957 and 1964 (NLSY79). The coefficients in Table 6.3 and the standard deviations of weight in pounds in Table 6.A3 imply that for two identical white women who differ in weight by two standard deviations (roughly 75 pounds), we would expect the lighter one to earn 3.7 percent higher wages than the heavier one. The effect for the NLSY79 cohort is 8.7 percent. There is a difference which is roughly 5.5 percent between the two coefficients.

The bias against weight appears to have decreased. None of the coefficients of weight in pounds are significant for Black and Hispanic females born between 1980 and 1984 (NLSY97). The negative effects of obesity on wages of white males also disappear in the more recent data. When using weight lagged two years for white females, the coefficients are smaller for all the measures of weight for the NLSY97 cohort compared to the NLSY79 cohort.

This research provides an initial attempt at determining whether or not there have been any major changes over time in the relationship between wages and weight. More research is needed, and as we get more years of data from NLSY97, we should be able to say more about whether or not the changes noted here hold as the NLSY97 cohort ages. If the differences noted in this chapter are real, what are some possible explanations for the change? For example are people in the 21st century more likely to accept overweight and/or obese people in the workplace? Are white women more comfortable with their weight, whatever it is? Do younger white women employees have higher self esteem in spite of their excess weight? Or it could be something else?

Table 6.1
Coefficients and t-Statistics from Log Wage Regressions for Females Aged 19-29 between 1981-1986

	White Female			Black Female			Hispanic Female		
	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects
Column Number	1	2	3	4	5	6	7	8	9
BMI	-.0088	-.0074	-.0053	-.0036	-.0024	-.0006	-.0029	-.0016	-.0035
	(-6.12)	(-3.11)	(-2.05)	(-2.35)	(-0.97)	(-0.22)	(-1.23)	(-0.51)	(-0.78)
Weight in pounds	-.0014	-.0012	-.0008	-.0005	-.0003	-.00001	-.0005	-.0003	-.0007
	(-6.03)	(-3.11)	(-2.02)	(-2.25)	(-0.72)	(-0.03)	(-1.29)	(-0.55)	(-0.89)
Underweight	.0395	.0279	-.0173	-.0552	-.129	-.1023	-.0353	-.0729	.0271
	(1.74)	(0.56)	(-0.65)	(-1.54)	(-1.55)	(-2.02)	(-0.81)	(-0.85)	(0.56)
Overweight	-.065	-.0312	-.0453	-.0383	-.019	-.00624	-.0155	-.0298	-.061
	(-4.26)	(-1.06)	(-2.31)	(-1.95)	(-0.44)	(-0.25)	(-0.65)	(-0.64)	(-2.16)
Obese	-.1176	-.1195	-.066	-.0627	-.0644	.0045	-.0796	-.0008	-.1726
	(-5.01)	(-3.27)	(-2.22)	(-2.5)	(-1.68)	(0.13)	(-2.27)	(-0.02)	(-3.2)
Number of observations	11941	2798	11941	3949	1025	3949	2782	690	2782

Notes:
1) Data: NLSY79 females aged between 16-29
2) One of three measures of weight is used: BMI, weight in pounds (controlling for height in inches) or indicator variables for clinical weight classification: underweight, overweight, and obese (where healthy weight is the excluded categories).
3) For BMI and weight in pounds, coefficients and t-statistics are listed. For indicators of clinical weight classification, the percent change in log wages associated with a change in the indicator variables from 0 to 1 and t-statistics are listed.
4) Other regressors include: number of children ever born, age of youngest child, general intelligence, highest grade completed, mother's highest grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, county unemployment rate, current school enrollment, part time job, white collar job, and region of residence.

Table 6.2
Coeficient and t-Statistic from Log Wage Regressions for Males Aged 19-29 between 1981-1986

	White Male			Black Male			Hispanic Male		
	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects
Column Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
BMI	.0001	-.0045	-.0034	.0002	.0048	.0004	-.0041	-.0058	-.0054
	(0.07)	(-1.63)	(-1.02)	(0.07)	(1.18)	(0.11)	(-1.44)	(-1.20)	(-1.10)
Weight in pounds	0.0000	-.0006	-.0004	.00001	.0005	.0001	-.0007	-.0009	-.0013
	(0.00)	(-1.65)	(-0.89)	(0.04)	(0.85)	(0.18)	(-1.62)	(-1.27)	(-1.91)
Underweight	-.1480	-.2408	-.0313	-.0152	-.0743	.1509	.1592	.4121	.2189
	(-3.53)	(-2.95)	(-0.70)	(-0.25)	(-0.74)	(2.32)	(1.42)	(2.41)	(1.36)
Overweight	.0289	.0042	.0177	.0128	.0140	.0024	-.0118	.0142	-.0273
	(2.01)	(0.18)	(1.16)	(0.53)	(0.36)	(0.09)	(-0.45)	(0.33)	(-1.00)
Obese	-.0434	-.1180	-.0004	.0127	.0709	.0010	-.0145	-.0497	.0027
	(-1.68)	(-3.60)	(-0.02)	(0.36)	(1.43)	(0.02)	(-0.36)	(-0.81)	(0.04)
Number of observations	12998	3399	33735	4335	1185	16025	3275	880	3275

Notes:
1) Data: NLSY79 males
2) One of three measures of weight is used: BMI, weight in pounds (controlling for height in inches) or indicator variables for clinical weight classification: underweight, overweight, and obese (where healthy weight is the excluded categories).
3) For BMI and weight in pounds, coefficients and t-tatistics are listed. For indicators of clinical weight classification, the percent change in log wages associated with a change in the indicator variables from 0 to 1 and t-statistics are listed.
4) Other regressors include: number of children ever born, age of youngest child, general intelligence, highest grade completed, mother's highest grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, county unemployment rate, current school enrollment, part time job, white collar job, and region of residence.

Table 6.3
Coefficient and t-Statistic from Log Wage Regressions for Females Aged 16-26 between 2001-2008

	White Female			Black Female			Hispanic Female		
	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects
Column Number	1	2	3	4	5	6	7	8	9
BMI	-.0030	-.0071	.0031	-.0029	-.0025	.0014	.0068	.0076	-.0091
	(-1.86)	(-3.40)	(0.67)	(-1.58)	(-1.14)	(0.23)	(1.86)	(1.35)	(-1.04)
Weight in pounds	-.0005	-.0011	.0005	-.0005	-.0004	.0003	.0013	.0015	-.0017
	(-1.92)	(-3.33)	(0.71)	(-1.67)	(-1.05)	(0.34)	(1.92)	(1.47)	(-1.10)
Underweight	.0156	.0328	.1072	.0831	-.0075	-.0306	-.1926	-.1005	.0148
	(0.29)	(0.42)	(1.44)	(1.41)	(-0.10)	(-0.40)	(-2.31)	(-0.94)	(0.11)
Overweight	-.0222	-.0849	.0088	.0231	-.1230	.0940	.0367	.0460	-.0448
	(-0.96)	(-2.59)	(0.25)	(0.62)	(-2.04)	(1.66)	(0.85)	(0.73)	(-0.72)
Obese	-.0413	-.0949	.0578	-.0216	-.0453	.1231	.0650	.0731	-.0894
	(-1.51)	(-2.76)	(1.00)	(-0.58)	(-1.07)	(1.22)	(1.39)	(0.97)	(-1.01)
Number of observations	3922	2248	3922	1535	807	6300	1067	544	1067

Notes:
1) Data: NLSY97 females aged 16-28
2) One of three measures of weight is used: BMI, weight in pounds (controlling for height in inches) or indicator variables for clinical weight classification: underweight, overweight, and obese (where healthy weight is the excluded categories).
3) For BMI and weight in pounds, coefficients and t-statistics are listed. For indicators of clinical weight classification, the percent change in log wages associated with a change in the indicator variables from 0 to 1 and t-statistics are listed
4) Other regressors include: number of children ever born, general intelligence, highest grade completed, mother's highest grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, current school enrollment, part time job, white collar job, and region of residence.

Table 6.4
Coeficient and t-Statistic from Log Wage Regressions for Males Aged 16-29 between 1981-2006

	White Male			Black Male			Hispanic Male		
	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects	OLS	OLS with Lagged Weight	Fixed Effects
Column Number	1	2	3	4	5	6	7	8	9
BMI	.0026	.0005	.0069	-.0036	-.0081	-.0061	-.0025	-.0008	.0070
	(1.17)	(0.18)	(1.36)	(-1.15)	(-2.52)	(-0.85)	(-0.77)	(-0.19)	(0.76)
Weight in pounds	.0003	.00007	.0013	-.0004	.0001	-.0007	-.0004	-.0002	.0014
	(1.14)	(0.19)	(1.74)	(-1.15)	(0.39)	(-0.72)	(-1.05)	(-0.33)	(0.96)
Underweight	.0144	.0353	.0427	-.1068	.1166	-.0576	-.1783	.0726	-.2251
	(0.26)	(0.34)	(0.49)	(-2.09)	(2.03)	(-0.50)	(-2.78)	(0.76)	(-1.53)
Overweight	.0020	.0085	.0004	.0035	-.0370	.0093	.0320	.0356	.0421
	(0.09)	(0.29)	(0.01)	(0.09)	(-0.75)	(0.14)	(0.71)	(0.70)	(0.62)
Obese	.0252	.0261	.0206	-.0471	.0326	-.0647	-.0192	.0144	.1100
	(0.81)	(0.64)	(0.32)	(-1.04)	(0.59)	(-0.66)	(-0.44)	(0.23)	(1.25)
Number of observations	4684	2887	4684	1452	878	1452	1375	812	1375

Notes:
1) Data: NLSY97 males aged 16-29
2) One of three measures of weight is used: BMI, weight in pounds (controlling for height in inches) or indicator variables for clinical weight classification: underweight, overweight, and obese (where healthy weight is the excluded categories).
3) For BMI and weight in pounds, coefficients and t-statistics are listed. For indicators of clinical weight classification, the percent change in log wages associated with a change in the indicator variables from 0 to 1 and t-statistics are listed.
4) Other regressors include: number of children ever born, general intelligence, highest grade completed, mother's highest grade completed, father's highest grade completed, years of actual work experience, job tenure, age, year, and indicator variables for marital status, current school enrollment, part time job, white collar job, and region of residence.

Table 6.A1
Summary Statistics for Females NLSY79 Aged 16-29

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Log wage	18672	6.21	0.52	4.24	10.60
Body mass index	18672	23.99	5.02	12.28	59.01
Height in inches corrected	18672	64.30	2.38	51.82	74.44
Weight in ounds corrected	18672	141.25	31.02	81.38	354.59
2-year lag BMI	4513	24.25	5.17	14.69	57.99
2-year lag weight	4513	142.65	31.96	81.52	343.93
General intelligence	18672	44.34	26.81	1	99
Highest grade completed	18672	12.95	2.09	0	20
Mother's higest grade completed	18672	11.18	2.99	0	20
Father's highest grade completed	18672	11.19	3.79	0	20
Year	18672	1986.04	3.42	1981	1994
Unemployment rate	18672	75.03	31.11	18	237
Number of weeks at current job	18672	104.72	115.37	1	724
Years of actual work experience	18672	5.28	3.21	0.02	14.76
Age	18672	24.32	3.37	16	29
Hispanic	18672	0.15	0.36	0	1
Black	18672	0.21	0.41	0	1
Underweight	18672	0.05	0.23	0	1
Normal weight	18672	0.63	0.48	0	1

Table 6.A1
Summary Statistics for Females NLSY79 Aged 16-29 (continued)

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Overweight	18672	0.20	0.40	0	1
Obese	18672	0.12	0.32	0	1
underw_2	4513	0.05	0.22	0	1
normalw_2	4513	0.61	0.49	0	1
Work more than 20 hours per week	18672	0.84	0.37	0	1
Blue collar job	18672	0.37	0.48	0	1
Married spouse present	18672	0.49	0.50	0	1
Never married	18672	0.39	0.49	0	1
Enrolled in school	18672	0.27	0.44	0	1
Not Enrolled in school	18672	0.73	0.44	0	1
Northeast region	18672	0.19	0.39	0	1
North Central region	18672	0.24	0.43	0	1
South region	18672	0.38	0.49	0	1
noChild	18672	0.63	0.48	0	1
fiveYChild	18672	0.24	0.42	0	1
tenYChild	18672	0.04	0.20	0	1
eighteenYChild	18672	0.00	0.06	0	1
adultYChild	18672	0.00	0.00	0	0

Table 6.A2
Summary Statistics for Males NLSY79 Aged 16-29

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Adjusted wage	20608	730.46	1322.12	69.20	42265.43
Log wage	20608	6.40	0.54	4.24	10.65
Body mass index	20608	24.74	4.02	12.33	61.03
Height in inches corrected	20608	69.73	2.63	60.55	79.74
Weight in pounds corrected	20608	171.44	31.08	87.17	412.21
2-year lag BMI	5464	25.09	3.98	14.52	50.80
2-year lag weight	5464	173.66	30.96	95.95	412.21
General intelligence	20608	42.03	29.38	1	99
Highest grade completed	20608	12.55	2.26	3	20
Mother's highest grade completed	20608	11.06	3.16	0	20
Father's highest grade completed	20608	11.06	3.90	0	20
Year	20608	1986.15	3.42	1981	1994
Unemployment rate	20608	75.18	31.31	13	237
Number of years at current job	20608	110.71	120.68	1	882
Years of actual work experience	20608	5.69	3.25	0.02	15.92
Age	20608	24.33	3.40	16	29
Hispanic	20608	0.16	0.37	0	1
Black	20608	0.21	0.41	0	1
Underweight	20608	0.02	0.13	0	1

Table 6.A2
Summary Statistics for Males NLSY79 Aged 16-29 (continued)

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Overweight	20608	0.30	0.46	0	1
Obese	20608	0.10	0.29	0	1
underw_2	5464	0.01	0.10	0	1
overw_2	5464	0.32	0.47	0	1
obese_2	5464	0.11	0.31	0	1
Work more than 20 hours per week	20608	0.90	0.30	0	1
White collar job	20608	0.33	0.47	0	1
Married spouse present	20608	0.59	0.49	0	1
Never married	20608	0.34	0.47	0	1
Enrolled in school	20608	0.24	0.43	0	1
Not Enrolled in school	20608	0.76	0.43	0	1
Northeast region	20608	0.19	0.39	0	1
North Central region	20608	0.26	0.44	0	1
South region	20608	0.35	0.48	0	1
noChild	20608	0.75	0.43	0	1
fiveYChild	20608	0.16	0.36	0	1
tenYChild	20608	0.01	0.11	0	1
eighteenYChild	20608	0.00	0.05	0	1
adultYChild	20608	0.00	0.01	0	1

Table 6.A3
Summary Statistics for Females NLSY97 Aged 16-29

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Adjusted wage	6527	528.33	663.01	46.44	26469.03
Log wage	6527	6.08	0.57	3.84	10.18
Body mass index	6527	26.29	6.53	14.06	67.27
Height in inches corrected	6527	64.61	2.50	51.90	80.34
Weight in pounds corrected	6527	156.33	40.78	78.87	372.44
2-year lag BMI	3599	25.93	6.31	15.68	59.45
2-year lag weight	3599	154.26	39.67	81.51	366.43
General intelligence	6527	52030.84	28174.84	0	100000
Highest grade completed	6527	2.05	1.17	0	7
Mother's highest grade completed	6527	13.09	2.83	1	20
Father's highest grade completed	6527	13.06	3.66	2	95
Year	6527	2004.06	2.27	2001	2008
Number of years at current job	6527	0.52	0.60	0	12.5
Years of actual work experience	6527	4.04	2.69	0	19.28
Age	6527	21.86	2.66	16	29
Hispanic	6527	0.16	0.37	0	1
Black	6527	0.24	0.42	0	1
Underweight	6527	0.03	0.17	0	1
Overweight	6527	0.24	0.43	0	1

Table 6.A3
Summary Statistics for Females NLSY97 Aged 16-29 (continued)

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Obese	6527	0.22	0.41	0	1
underw_2	3599	0.03	0.18	0	1
overw_2	3599	0.23	0.42	0	1
obese_2	3599	0.20	0.40	0	1
Work more than 20 hours per week	6527	0.71	0.46	0	1
Work less than 20 hours per week	6527	0.29	0.46	0	1
White collar job	6527	0.27	0.44	0	1
Married spouse present	6527	0.88	0.33	0	1
Never married	6527	0.11	0.31	0	1
Married spouse not present	6527	0.02	0.13	0	1
Enrolled in school	6527	0.40	0.49	0	1
Northeast region	6527	0.16	0.37	0	1
North Central region	6527	0.23	0.42	0	1
South region	6527	0.39	0.49	0	1
West region	6527	0.22	0.42	0	1

Table 6.A4
Summary Statistics for Males NLSY97 Aged 16-29

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Adjusted wage	7517	631.03	1205.60	46.44	27793.22
Log wage	7517	6.18	0.61	3.84	10.23
Body mass index	7517	25.92	5.27	13.90	56.73
Height in inches corrected	7517	70.18	2.71	56.64	82.75
Weight in pounds corrected	7517	181.87	39.85	102.82	499.12
2-year lag BMI	4577	25.53	5.13	15.27	54.66
2-year lag weight	4577	178.98	38.71	106.02	446.62
General intelligence	7517	47366.62	29904.90	0	100000
Highest grade completed	7517	1.77	1.11	0	7
Mother's highest grade completed	7517	12.99	3.11	1	95
Father's highest grade completed	7517	12.92	4.58	2	95
Year	7517	2004.12	2.26	2001	2008
Number of years at current job	7517	0.53	0.61	0	8.34
Years of actual work experience	7517	4.01	2.68	0	17.88
Age	7517	21.84	2.61	16	29
Number of kids	7517	0.15	0.48	0	6
Hispanic	7517	0.18	0.39	0	1
Black	7517	0.19	0.39	0	1
Underweight	7517	0.01	0.12	0	1

Table 6.A4
Summary Statistics for Males NLSY97 Aged 16-29 (continued)

Variable	N	Mean	Standard Deviation	Minimum	Maximum
Overweight	7517	0.30	0.46	0	1
Obese	7517	0.18	0.39	0	1
underw_2	4577	0.01	0.12	0	1
overw_2	4577	0.29	0.45	0	1
obese_2	4577	0.16	0.37	0	1
Work more than 20 hours per week	7517	0.80	0.40	0	1
White collar job	7517	0.17	0.38	0	1
Married spouse present	7517	0.92	0.28	0	1
Never married	7517	0.07	0.26	0	1
Married spouse not present	7517	0.01	0.10	0	1
Enrolled in school	7517	0.32	0.46	0	1
Not Enrolled in school	7517	0.68	0.46	0	1
Northeast region	7517	0.16	0.37	0	1
North Central region	7517	0.25	0.44	0	1
South region	7517	0.35	0.48	0	1

Chapter 7

CONCLUSION

The recent increase in obesity, with its negative economic, social and health consequences has attracted considerable attention from applied economists. As a result, I investigate the relationship between body weight and wages in the U.S. by gender, race, and ethnicity. I first replicate Cawley (2004) with the same data (NLSY79 1979-2000), variables, and methods. My results are similar to Cawley's ordinary least squares estimation which indicate that heavier white females, Black females, Hispanic females and Hispanic males tend to earn less, and heavier Black males tend to earn more, than their lighter counterparts. In order to avoid the influence of wages on current weight, I estimated models using lagged body weight. One consistent result is that weight appears to lower the wages of white females; this finding is consistent for OLS with current weight and OLS using lagged weight. Individual fixed effects estimations eliminate the time-invariant unobserved heterogeneous effects on wages and has the dramatic effect of eliminating the negative correlation between BMI and weight in pounds and wages for all groups but white females. OLS estimates for white women indicate that a difference in weight of two standard deviations (roughly 65 pounds) is associated with a difference in wages of nine percent. In absolute value, this is equivalent to the effect of roughly one and half years of education or three years of work experience. The negative relationship between weight and wages observed for other gender-ethnic groups appears to be due to unobserved heterogeneity.

I also estimated the same models as Cawley's (methods and variables) to check the effect of weight on wages but with a larger data set which is extended by three years (NLSY79 1981-2006) and again found similar results to those found by Cawley (2004).

I contributed to the existing literature by providing new evidence on the relation between weight and wages by asking the question "Does the way missing data is handled cause different results from Cawley." His approach utilizes dummy variables for each variable to flag missing data when it occurs. My approach deletes an observation whenever there is missing data for a particular explanatory variable and is called censored regressors (Rigobon and Stoker (2009)). We are unaware of studies of the effects of weight on wages using this approach for treating missing data. I used data for 1981 through 2000 from the NLSY79 data set to compare the results with those of Cawley. Then I added the 2002, 2004 and 2006 data. When using this approach, we find, for both data sets used, that ordinary least squares and individual fixed effects lead to results that are similar to those of Cawley (2004) for all sub-groups but Black males. For the last group, none of the coefficients were statistically significant when using censored regressors. The hypotheses that the same coefficients (BMI, lagged-7 BMI and lagged-7 weight for Black males) from the two different regressions(Censored regressors and Cawley's replication) are significantly different, were rejected at the 10 percent significance level.

Moreover none of the coefficients for white males where significant when using censored regressors. One exception, when using censored regressors, is for white males who were obese seven years earlier. The coefficients in Table 4.2 and the standard deviations in Table A4.2 imply that for two otherwise identical white males, one with a mean weight and one at the two standard deviations above the mean, we would expect the lighter one to enjoy 9 percent higher wages. This wage difference de roughly equal in magnitude to the difference associated with 1.5 more years of education, or three more years of work experience.

Using data for the period 1981 through 2006, quantile regressions helped to detect if the effects of weight on wages are the same throughout the distribution of wages across gender, race, and ethnicity. I find some evidence that the weight penalty on wages increases with wages for almost all sub-groups but Black males.

The results for the quantile regression are consistent with the notion that these women don't let weight affect their choice of an occupation paying higher wages even though the weight-wage penalty appears to get larger as wages rise.

The sociological literature yields a possible explanation for result that the weight penalty is greater for white females than it is the case for Black and Hispanic females. The explanation is that obesity has a more adverse impact on the self-esteem of white females than on that of Black and Hispanic females, who report perceiving higher weight as a signal of power and stability (Stearns 1997). More research is needed to explain differences across gender and race and to explain the increasing penalty associated with higher wages. Is it due to discrimination by employers regarding physical appearance? Is it due to the low self-esteem of the employees with excess weight? Or is it something else?

Finally chapter six investigates the changes of the effects of weight on wages in the U.S. over time, across gender and race. For this purpose, I compare the effects of weight on wages between two generations aged 16 to 29. The first generation, born between 1957 and 1964, is taken from the National Longitudinal Survey of Youth 1979 (NLSY79). The second cohort is taken from the National Longitudinal Survey of Youth 1997 (NLSY97) where respondents were born between 1980 and 1989.

Ordinary least squares results indicate that heavier females including white, Black, and Hispanic, born between 1957 and 1964, tend to earn less than their lighter counterparts. Models also are estimated using lagged body weight, in order to account for the possible influence of wages on contemporaneous weight. The results for the cohort based on the NLSY79 data in chapter 4 are similar to the results for the young cohort based on the NLSY79. For example, if we take two identical white females born between 1957 and 1964, aged between 16 and 29 and who differ in weight by two standard deviations (roughly 58 pounds), we would expect the lighter one to earn wages that are 8.12 percent higher.

However, the results are different for females born between 1980 and 1984 since the weight penalty appears to have decreased. None of the coefficients of weight in pounds variable are significant for Black and Hispanic females in the NLSY97. When using weight lagged two years for white females, the coefficients are smaller for all the measures of weight for the NLSY97 cohort compared to the NLSY79 cohort.

My results imply that for two identical white women who differ in weight by two standard deviations (roughly 75 pounds), we would expect the lighter one to earn 3.7 percent higher wages than the heavier one. The effect for the NLSY79 cohort is 8.7 percent. These results suggest that the effect of weight on wages has decreased for the white women of the 21st century.

This research provides an initial attempt at determining whether or not there have been any major changes over time in the relationship between wages and weight. More research is needed, and as we get more years of data from NLSY97, we should be able to say more about whether or not the changes noted here hold as the NLSY97 cohort ages. If the reduction in the weight wage penalty, suggested by the results from Chapter 6, are real what are some possible explanations for the change? For example are people in the 21st century more likely to accept overweight and/or obese people in the workplace? Are white women more comfortable with their weight, whatever it is? Do younger white women employees have higher self esteem in spite of their excess weight? Or it could be something else? Obesity is preventable. Thus, public health policies should be developed and implemented with the object of decreasing the incidence of obesity and thus its negative effect on wages and on health.

References

Averett, Susan, and Sanders Korenman. 1996. "The Economic Reality of the Beauty Myth" *Journal of Human Resources* 31(2):304-30

Averett, Susan, and Sanders Korenman. 1999, " Black-White differences in Social and Economic Consequences of Obesity" *International Journal of Obesity* 23:2 (February), 166-173.

Banner, Lois. 1983." American Beauty". Chicago, US: University of Chicago Press.

Baum, Charles L. II and William F. Ford. 2004. "The wage effects of obesity: a Longitudinal study." *Health Economics* 13: 885-89.

Behrman, Jere R., and Mark R. Rosenzweig. 2001. "The Returns to Increasing Body Weight." Unpublished.

Bhattacharya, Jay and M. Kate Bundorf. April 2005. "The Incidence of the Healthcare Costs of Obesity." *National Bureau of Economic Research* NBER Working Paper No. 11303.

Brunello, Giorgio & D'Hombres, Beatrice, 2007. "Does body weight affect wages?: Evidence from Europe," *Economics and Human Biology, Elsevier,* vol. 5(1), pages 1-19, March.

Burkhauser, Richard and John Cawley. 2002. "Obesity, Disability, and Movement onto The Disability Insurance Rolls." University of Michigan Retirement Research Center Working Paper No. 089.

Burkhauser, Richard V. & Cawley, John, 2008. "Beyond BMI: The value of more accurate measures of fatness and obesity in social science research," *Journal of Health Economics*, Elsevier, vol. 27(2), pages 519-529, March.

Cawley, J. and S. Danzinger. 2004. "Obesity as a Barrier to the Transition from Welfare to Work." *National Bureau of Economic Research*, working paper # 10508, May.

Cawley, J., Grabkal, M.M., Lillard, D.R., 2005. "A comparison of the relationships between obesity and earnings in the US and Germany." *Journal of Applied Social Science Studies* 125 (1), 119–129.

Cawley, J., S. Markowitz and J. Tauras. 2004. "Lighting Up and Slimming Down: The Effects of Body Weight and Cigarette Prices on Adolescent Smoking Initiation." *Journal of Health Eonomics* 23:293-311.

Cawley, John and C. Katharina Spiess. "Obesity and Skill Attainment in Early Childhood." *Economics and Human Biology*, 2008, 6(3): 388-397.

Cawley, John, Sara Markowitz, and John Tauras. "Lighting Up and Slimming Down: The Effects of Body Weight and Cigarette Prices on Adolescent Smoking Initiation." *Journal of Health Economics*, 2004, 23(2): 293-311.

Cawley, John. "The Economics of Childhood Obesity." *Health Affairs*, 2010, 29(3): 364-371.

Cawley, John. 2000. “Body Weight and Women’s Labor Market Outcomes.” *National Bureau of Economic Research* NBER Working Paper No. 7841.

Cawley, John. 2004. “The Impact of Obesity of Wages” *Journal of Human Resources*, 39(2):451-74

Chou, S. Y., M. Grossman and H. Saffer. 2004. "An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System." *Journal of Health Economics* 23: 565-587.

Coestier, B., E. Gozlan and S. Marette. 2005. "On Food Companies Liability for Obesity."*American Journal of Agricultural Economics* 87: 1-14.

Conley, Dalton and Glauber, Rebecca. 2008.”Gender, Body Mass and Economic Statu.” *National Bureau of Economic Research NBER* Working Paper No. W11343. Available at SSRN: http://ssrn.com/abstract=727123

Cox, B.D., Whichelow, M. (1996), *"*Ratio of waist circumference to height is better predictor of death than body mass index*"*, *British Medical Journal Bmj*, Vol. 313 No.7070, pp.1487.

Dalton M, Cameron AJ, Zimmet PZ et al.’’Waist circumference, waist–hip ratio and body mass index and their correlation with cardiovascular disease risk factors in Australian adults.’’ *Journal of Internal Medicine* 2003; **254**: 555–63.

De Garine, I. 1993. "Obesity in Multicultural Perspective." In *Food, Diet and Economic Change Past and Present,* edited C. Geissler and D.J. Oddy. New York: St. Martin's Press, 175-182.

Ezzati, M., Martin, H., Skjold, S., Vander Hoorn, S., Murray, C. "Trends in national and state-level obesity in the USA after correction for self-report bias analysis of health surveys" *Journal of the Royal Society of Med*icine 99 (5), 2006, 250–257.

Finkelstein, E. Justin G. Trogdon, Joel W. Cohen and William Dietz "Annual Medical Spending Attributable To Obesity: Payer-And Service-Specific Estimates" *Health Affairs*, 28, no. 5 (2009): w822-w831 (Published online 27 July 2009) doi: 10.1377/hlthaff.28.5.w822

Flegal, K. M., M. D. Carroll, C. L. Ogden and C. L. Johnson. 2002. "Prevalence and Trends in Obesity Among US Adults." *Journal of American Medical Association* 288: 1723-1727.

Flegal, KM, et. al. 2002. "Prevalence and trends in obesity among US adults, 1999-2000." *The Journal of the American Medical Association,* 288(14) (October):1772-3.

Gortmaker, S., Must, A., Perrin, J., Sobol, A., Dietz, W., 1993. "Social and economic consequences of overweight in adolescence and young adulthood." . *The New England Journal of Medicine.* 329, 1008–1012.

Hamermesh, D. S. and J. E. Biddle. 1994. "Beauty and the Labor Market." *American Economic Review* 84:1174-1194.

Han E, Norton EC, Stearns SS. "Weight and wages: Fat versus lean paychecks". *Health Economics*. 18(5) 535-548 (2009)

Han, Euna; Norton, Edward C.; Powell, Lisa M. 2009. "Direct and Indirect Effects of Teenage Body Weight on Adult Wages" *National Bureau of Economic Research, Inc,* NBER Working Papers: 15027 AN: 1045928

Haskins, K. and H. E. Ransford (1999). "The Relationship Between Weight and Career Payoffs Among Women." *Sociological Forum.* Vol. 14 (June, 1999), pp. 295-318.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica.* January,47:1, pp. 153–61.

Ippolito, P. M. and A. Mathios. 1995. "Information and Advertising: The Case of Fat Consumption in the United States." *American Economic Review,* AEA Papers and Proceedings 85: 91-95.

Kan, K. and W. D. Tsai. 2004. "Obesity and Risk Knowledge." *Journal of Health Economics* 23: 907-934.

Koenker, R., & Hallock, K. (2001). Quantile regression [Electronic version]. *Journal of Economic Perspectives, 15*(4), 143-156.

Lempert David, 2007. "Women's Increasing Wage Penalties from Being Overweight and Obese," Working Papers 414, U.S. *Bureau of Labor Statistics.*

Loh, Eng Seng. 1993. "The Economic Effects of Physical Appearance."*Social Science Quarterly* 74(2):420-38.

Lundborg, P. et al." Obesity and occupational attainment among the 50+ of Europe"2007.

Mavromara, K . 2008. "Obesity as a Multidisciplinary Research Topic: Editor Introduction" *Australian Economic Review*, **41**, (1), 67-71

Mobius, Markus and Tanya Rosenblat. 2006. "Why Beauty Matters?" *American Economic Review,* 96: 222-35.

Nayga, R. M. 2000. "Schooling, Health Knowledge and Obesity." *Applied Economics* 7: 815-822.

Nayga, R. M. 2001. "Effects of Schooling on Obesity: Is Health Knowledge a Moderating Factor?" *Education Economics* 9:129-137.

Ossiander, E., Emanuel, M., O'Brien, I., Malone, W.K."Drivers licenses as a source of data on height and weight" *Economics and Human Biology*. 2 (2), 2004, 219–227.

Pagan, Jose, and Alberto Davila. 1997. "Obesity, Occupational Attainment, and Earnings." *Social Science Quarterly* 78:3 (September), 756-70.

Philipson, T. 2001."The World-Wide Growth in Obesity: An Economic Research Agenda." *Health Economics* 10: 1-7.

Plankey, MW, J Stevens, KM Flegal, and PF Rust, 1997. "Prediction Equation do not Eliminate Systematic Error in Self-Reported Body Mass Index." *Obesity Research*, 5:308-314.
Register, Charles A. and Donald R. Williams. 1990. "Wage Effects of Obesity Among Young Workers." *Social Science Quarterly* 71(1):130-41.

Roberto Rigobon, Thomas M. Stoker 2009." Bias From Censored Regressors*"* *Journal of Business and Economic Statistics*. July 1, 2009, 27(3): 340-353. doi:10.1198/jbes.2009.06119.

Ruhm, Christopher J. 2007 "Current and Future Prevalence of Obesity and Severe Obesity in the United States," *Forum for Health Economics & Policy*: Vol. 10: Iss. 2 (Obesity), Article 6 http://www.bepress.com/fhep/10/2/6

Sargent, J., Blanchflower, D., 1994. "Obesity and stature in adolescence and earnings in young adulthood". *Analysis of a British birth cohort. Arch. Pediat. Adolesc*. Med. 148, 681–687.

Shen, Wei, et. al. 2006. "Waist Circumference Correlates with Metabolic Syndrome Indicators Better Than Percentage Fat." *Obesity* 14(4):727-36.

Sönmez K; Akçakoyun M; Akçay A; Demir D; Duran N E; Gençbay M; Degertekin M; Turan F.
"Which method should be used to determine the obesity, in patients with coronary artery disease? (body mass index, waist circumference or waist-hip ratio).'' *International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity* 2003;27(3):341-6.

Stoker, Thomas M. and Rigobon, Roberto, 2007 " Estimation with Censored Regressors: Basic Issues". *International Economic Review*, Vol. 48, Issue 4, pp. 1441-1467, November 2007. Available at SSRN: http://ssrn.com/abstract=1071239 or doi:10.1111/j.1468-2354.2007.00470.x

Sturm, R. 2002." The Effects of Obesity, Smoking, and Drinking on Medical Problems and Costs" *Health Affairs* ; 21; 245-253.

Thorpe KE, Florence CS, Howard DH, Joski P. 2004 "Trends; The Impact of Obesity on Rising Medical Spending." *Health Affairs, Web Exclusive*, October 20, 2004.