The Dynamic Effects of Obesity on the Wages of Young Workers*

Joshua C. Pinkston
University of Louisville

March 23, 2017

Abstract

This paper considers effects of body mass on wages in the years following labor market entry. The preferred models allow current wages to be affected by both past and current body mass, as well as past wages, while also addressing the endogeneity of body mass. I find that a history of severe obesity has a large negative effect on the wages of white men. White women face a penalty for a history of being overweight, with some evidence of additional penalties that begin above the threshold for severe obesity. Furthermore, the effects of past wages on current wages imply that past body mass has additional, indirect effects on wages, especially for white women.

*I would like to thank Chuck Courtemanche, Jose Fernandez, Antonio Galvao and Christina Pinkston for helpful comments. I would also like to thank seminar and conference participants at the University of Tokyo; the University of Wisconsin, Milwaukee; Western Kentucky University; the 2011 IHEA World Congress; the 2011 Southern Economics Association Meetings; and the 2013 Society of Labor Economists Meetings.
Email: josh.pinkston@louisville.edu.

This paper considers effects of past and current body mass on wages in the early years of workers' careers. Using data from the National Longitudinal Survey of Youth 1997, I estimate dynamic models of wages in which body mass is allowed to be endogenous. This approach allows workers' history in the labor market, including past wages, to affect current wages. As a result, effects of high body mass can accumulate and persist over time. Furthermore, the preferred estimates are unbiased even when body mass is correlated with both individual fixed effects and time-varying unobservables.

The literature on body mass and wages has been understandably concerned with the endogeneity of body mass. Cawley (2004) carefully describes the reasons weight or body mass may be correlated with fixed and time-varying unobserved heterogeneity; however, most previous work has focused on either individual fixed effects or time-varying sources of endogeneity, but not both in the same regression.[1]

One contribution of this paper is that the estimation addresses multiple sources of bias simultaneously. The use of autoregressive wage equations eliminates a source of omitted variable bias that has been ignored by previous work, and individual fixed effects are removed by first-differencing. The panel data are then exploited to address endogeneity associated with remaining,

[1]E.g., Han et al. (2011) use fixed effects. Baum and Ford (2004) use first differences. Kline and Tobias (2008) use parent's BMI as an instrument. Cawley (2004) and Sabia and Rees (2012) use fixed effects and IV in separate regressions.

Two exceptions, Averett and Korenmann (1996) and Behrman and Rosenzweig (2001), use differencing to remove sibling (or twin) fixed effects and then use lagged BMI as either a proxy or an instrument for current BMI.

time-varying errors. Furthermore, I submit my identifying assumptions to considerable scrutiny, and find that they survive a variety of tests.

Another important contribution of this paper is the estimation of models that are consistent with discrimination affecting labor market search, occupational sorting or other channels that imply dynamic effects of body mass on wages. The preferred specifications allow both current and past body mass to affect wages. Furthermore, wages are affected not only by the lags of body mass included in the model, but also by further lags of body mass that have indirect effects through their effects on lagged wages.

In contrast, previous work has assumed that wage penalties associated with obesity are the same whether the worker recently became obese or had been obese her entire career. An obvious exception is a recent paper by Chen (2012), which examines the effects of current BMI and BMI 10 years earlier on the wages of workers in their 30s.[2] The current paper provides a detailed look inside of those early years using a more recent cohort of workers.[3]

This paper is also the first in the literature to use the NLSY97, and one of the first to focus on workers who entered the labor market more recently than the 1980s.[4] Odgen et al. (2010) report that obesity doubled among

[2]Han et al. (2011) consider direct effects of current body mass on wages along with indirect effects due to the influence of adolescent BMI on education and occupation.

[3]Another difference between Chen (2012) and the current paper is the treatment of endogeneity. Chen (2012) notes that adding future body mass to her regressions suggests a problem with endogeneity, but does not otherwise address the problem.

[4]Sabia and Rees (2012) replicate Cawley (2004) using a sample from Add Health that was between 24 and 32 in 2008; however, the design of Add Health limits their ability to perform the FE or IV estimation of Cawley (2004).

adults in the U.S. between 1980 and 2000. If prejudice or stereotypes evolve as the population becomes heavier, results from previous work based on the NLSY79 may not generalize to more recent cohorts.[5]

Finally, the estimation sample is unique in its focus on young workers in the first several years after entry into the labor market. Wage growth is higher earlier in careers, job and occupational mobility are more important, and negative shocks early in a career have lasting effects.[6] As a result, estimation with a sample of younger workers should be better able to capture potential discrimination as it unfolds than estimation with an older sample would. Removing fixed effects in a sample of older workers is likely to remove the accumulated effects of past discrimination suggested by Chen (2012). Any signals inferred from body mass should also have larger effects for younger than for older workers because the market knows less about younger workers.[7]

The empirical results suggest that wages are affected by past body mass and past wages. White men are penalized for a history of severe obesity. White women face a penalty for being overweight at all in previous years, with some specifications suggesting additional penalties as past or current

[5] Altonji et al. (2012) find that the mix of skills and family backgrounds changed between NLSY cohorts, which would also affect the ability to generalize across decades.

[6] See Murphy and Welch (1992), Topel and Ward (1992), and Neal (1999) among others. Kahn (2010 and Oreopoulous et al. (2012) find that economic conditions at labor market entry have persistent effects.

[7] Hamermesh (2011) uses a similar argument to explain why effects of beauty on wages might decline with age. Altonji and Pierret (2001) discuss statistical discrimination with employer learning about worker productivity. Learning about healthcare costs may be more complicated than learning about productivity if the relationship between BMI and costs changes with age.

BMI exceeds the threshold for severe obesity. Furthermore, I find that including past wages in the model is critical for identification and has important implications for the interpretation of results, especially for women.

The next section discusses models of wages and body mass, building up to dynamic panel data specifications. Section 2 discusses the data. Section 3 describes tests of the identifying assumptions introduced in Section 1 before discussing other details of the estimation. Section 4 presents results, followed by various robustness tests. Section 5 concludes.

1 Empirical Models of Body Mass and Wages

Following the recent literature, our first attempt at specifying a wage regression to measure effects of body mass might take the form

$$w_{it} = X_{it}\beta + BMI_{it}\phi + \nu_{it}, \tag{1}$$

where w_{it} is the log of person i's wage in period t, X_{it} is a vector of observable characteristics, and BMI_{it} is a vector that describes body mass using dummy variables or a polynomial. BMI_{it} is potentially correlated with both fixed individual effects related to genetics or upbringing and time-varying factors in the error term, ν_{it}, leading to possible endogeneity.

Some previous work used fixed effects or first-differenced estimation to eliminate bias in equation (1), but potential correlation between BMI_{it} and time-varying shocks to wages then remains.[8] Other estimates used instrumental variables to address the bias in equation (1); however, the instruments used so far in the literature are correlated with individual fixed effects, and likely correlated with relevant time-invariant unobservables.[9] The two approaches could be combined, but doing so requires a valid instrument for changes in BMI_{it}.

An additional problem in the literature on weight and wages is that most authors have at least implicitly assumed that only current body mass affects wages.[10] This assumption is inconsistent with the broader discrimination literature. Models in which discrimination affects job search, hiring decisions or promotions would imply that a worker's history of body mass would affect her current wages.[11] Furthermore, any statistical discrimination story in which body mass is used as a signal of productivity or healthcare costs would imply effects of the entire history of body mass observed by the current employer.

[8] E.g., Cawley (2004), Baum and Ford (2004), and Han et al. (2009).

[9] The most common instrument in the previous literature is the BMI of a family member, which was first used by Cawley (2004); however, Han et al. (2009) and my own estimation suggest that sibling or parent's BMI predicts only time-invariant components of BMI.

More importantly, I find that siblings' BMI is correlated with AFQT scores in the NLSY79 and '97, even controlling for respondents' BMI. This is consistent with the warning in Cawley et al. (2011) that instruments based on genes may not be valid because a single gene can affect multiple relevant outcomes.

[10] As noted above, Chen (2012) and Han et al. (2011) are exceptions.

[11] E.g., Lang et al. (2005) show that effects of prejudice or expected productivity differences are magnified in a wage posting model, producing wage differences even if employers are not willing to pay more to hire their preferred group. Bjerk (2008) shows that statistical discrimination can result in wage differentials over time due to effects on hiring and promotions.

The common assumption that wages are affected only by current body mass is also inconsistent with the presence of labor-market frictions, regardless of how or why wage penalties arise. Employers that have difficulty lowering wages during economic downturns may have difficulty cutting wages in response to undesirable weight gain. Workers who face limited opportunities due to their weight may face difficulty moving to better jobs after losing weight, much as workers who enter the labor market during recessions face lower wages long after the economy recovers.[12]

If a history of being overweight or obese can affect current wages, regressions like equation (1) should be modified to allow effects of both current and past body mass. But body mass (past or current) may be affected by past wages, which are likely correlated with current wages. In addition to the potential simultaneity of wages and body mass that has been discussed in the literature, it is possible that BMI is predetermined by past wages. Therefore, lagged wages should be added to avoid omitted variable bias. Including lagged wages has the additional benefit of allowing lags of body mass that are not included in the regression to have indirect effects on current wages through their effects on lagged wages.

Using a single lag of both wage and body mass results in an autoregressive

[12]Kahn (2010) and Oreopoulos et al. (2012) discuss the persistent effects of entering the market during a recession. Jacobson et al. (1993) also find that plant closings and layoffs have persistent negative effects.

wage equation:

$$w_{it} = \gamma w_{it-1} + X_{it}\beta + BMI_{it}\phi + BMI_{it-1}\phi_1 + \alpha_i + \varepsilon_{it}. \qquad (2)$$

Both BMI_{it} and BMI_{it-1} are potentially correlated with the individual fixed effect, α_i, as is w_{it-1}.[13] As before, BMI_{it} may be correlated with the time-varying error, ε_{it}, or with earlier shocks to the wage.

Dynamic panel data models like equation (2) can be estimated using the differenced GMM estimator developed by Holtz-Eakin et al. (1988) and Arellano and Bond (1991) (HENR and AB in what follows).[14] The first step in this approach is to use differences to eliminate the fixed effect:

$$\Delta w_{it} = \gamma \Delta w_{it-1} + \Delta X_{it}\beta + \Delta BMI_{it}\phi + \Delta BMI_{it-1}\phi_1 + \Delta\varepsilon_{it}. \qquad (3)$$

After differencing, ΔBMI_{it} and ΔBMI_{it-1} may still be correlated with the error term, and Δw_{it-1} is correlated with $\Delta\varepsilon_{it}$ through ε_{it-1}.

Fortunately, further lagged levels of the wage are valid instruments for Δw_{it-1} if there is no serial correlation in ε. Under this assumption, w_{it-2} is not correlated with ε_{it} or ε_{it-1}, but is correlated with Δw_{it-1}.[15] The GMM estimator of HENR and AB also uses further lags, where available, as instruments to improve efficiency.

[13]Further lags of BMI or w can be included, but one lag is sufficient to explain the model. The lag structure is discussed further in Section 3.

[14]See Arellano and Honoré (2001), Bond (2002), or Arellano (2003) for reviews.

[15]The correlation of w_{it-2} and Δw_{it-1} is weak if γ is close to 1; however, the results presented below suggest this is not a problem.

Assuming no serial correlation in the time-varying errors, ε_{it}, is not equivalent to assuming no serial correlation in wages or wage growth. On the contrary, the autoregressive specifications of equations (2) and (3) assume that current wages are correlated with past wages, and current wage growth is correlated with past wage growth. The assumption of no serial correlation in ε is violated only if there is serial correlation in residual heterogeneity that is uncorrelated with lagged wages, BMI (lagged and current), and the regressors included in X_{it}.[16]

Lagged levels of BMI are valid instruments in the differenced estimator under an additional assumption. Specifically, BMI_{it-2} and further lags are valid instruments for ΔBMI_{it} and ΔBMI_{it-1} as long as BMI_{it} is uncorrelated with ε_{it+1} for all t. On an intuitive level, if the endogeneity of BMI_{it} is due to reverse causality, this assumption requires that random shocks to *future* wages do not affect *current* body mass. If endogeneity is due to unobserved shocks that are common to BMI_{it} and ε_{it}, this assumption requires that those shocks only affect w_{it+1} through their effects on BMI_{it} and w_{it}.

Finally, the identification of equation (3) requires changes in BMI to be predicted by its lagged levels. BMI_{it-2} should be correlated with ΔBMI_{it} if BMI_{it} is endogenous. This assumption finds empirical support in the dynamic models of body mass estimated by Goldman et al. (2010) and Ng et

[16]In contrast, serial correlation in the residuals of static wage regressions is expected because those residuals are not independent of lagged wages. The dynamic estimation in the current paper, therefore, supports the comments in Cawley (2004) about serial correlation in the previous literature.

al. (2010). Studies in the epidemiology literature also find that large changes in weight are more common among those who were initially heavier.[17] Lee et al. (2010) suggest that avoiding weight gain may require greater effort from overweight women than from normal weight peers. Finally, in supplemental regressions (available upon request) I find that BMI_{it-2} is correlated with ΔBMI_{it}, with F statistics above traditional cutoffs for weak instruments.[18]

2 Data

This paper uses data from the 1997 through 2009 waves of the National Longitudinal Survey of Youth 1997 (NLSY97). Individuals in the sample were between 12 and 16 years of age in 1996. They were between 24 and 30 when interviewed in 2009. The data also contain detailed information on labor market history, demographics, and other common control variables.

The NLSY97 has important advantages over the 1979 cohort for the purposes of this paper. The '97 respondents were young enough at their first interview that nearly all of them are observed as they begin their careers. They were also asked about height and weight in every year of the survey. In contrast, NLSY79 respondents were as old as 22 when first interviewed; and were not asked about weight in '79, '80, 83, '84 or '87. As a result, '79 cohort was between 25 and 33 years old in the first year (1990) that BMI

[17] Examples include Lewis et al. (2000) and Williamson et al. (1990).

[18] These regressions are not equivalent to the first stage of 2SLS. They are suggestive. I still consider the possibility that the instruments are too weak to identify coefficients on both BMI_{it} and BMI_{it-1} when I examine the robustness of my results.

could be observed for three years in a row.

In what follows, attention is limited to white men and women due to concern for sample sizes. Over twice as many respondents identify as white than as black, which is the second largest racial or ethnic group. Furthermore, requiring at least three consecutive years of wage observations reduces the sample size more for minorities and women than for white men.

The estimation sample only includes jobs that follow full-time labor market entry. I define entry as the first two consecutive years in which the individual works full time for at least 75% of the year. This restriction is intended to exclude the temporary or part-time jobs of younger workers that likely bear little resemblance to their eventual careers.[19] Part-time jobs that take place later in workers' careers are still included in the sample.

The sample also excludes respondents who were in the military, as well as observations for women who were pregnant at any point since the last interview. Outliers in the wage distribution are only dropped if $\Delta \ln(w_{it}) \geq 6.5$, leaving some observations with wages that may seem unusually high or low.[20] Limiting attention to observations that can be use as time t, $t-1$, or

[19]Nearly 75% of jobs excluded by the entry restriction are part-time, compared to 9% of jobs in the sample. Median tenure is 23 weeks for excluded jobs, but 85 weeks for jobs following full-time entry.

The sample includes people who entered the labor-market but returned to (or never left) school. The results discussed below are robust to excluding people who are in school, but some statistical significance is lost due to the smaller sample size.

[20]All of the results that follow are robust to the treatment of outliers in the wage distribution. Since these outliers are often the result of measurement error, robustness to the treatment of outliers suggests the results are at least not affected by the most severely mismeasured observations in the data.

$t-2$ in equation (3) leaves 9,037 observations for 1,473 white men and 5,408 observations for 1,060 white women.

A more detailed discussion of the sample's selection is left to an appendix. The rest of this section discusses information on body mass in the data, followed by a brief description of the estimation sample.

2.1 Body Mass and Measurement Error

The data include self-reported height and weight in each year, which allows the construction of BMI.[21] The measurement error introduced by the use of self-reported height and weight is well known and widely discussed; however, previous research has ignored the fact that roughly 10% of person/year observations in either NLSY cohort come from telephone interviews, which worsen misreporting relative to in-person interviews.[22] All regressions that include current or lagged BMI variables also include corresponding dummy variables for interviews being conducted by phone.[23]

I do not rescale self-reported height and weight based on the relationships

[21]BMI is defined as (weight in kg)/(height in m)^2

[22]White women are especially sensitive to interview method. Controlling for age and individual fixed effects, average reported weight falls by over 3.5 pounds when white women are interviewed by phone. Reported weight falls by over seven pounds when overweight women are interviewed by phone.

[23]The coefficients on indicators of phone interviews are uniformly small and statistically insignificant in estimated wage equations, suggesting that any measurement error in wages is at least less sensitive to interview mode than body mass is.

More generally, the literature on measurement error in wages suggests that the errors are mean-reverting. Furthermore, Pischke (1995) argues that the mean reversion is a combination of white noise and the magnitude of changes being understated. On an intuitive level, it seems the understatement of changes in income would work against this paper finding any association between BMI and changes in wages.

between reported and actual measures in NHANES data, as Cawley (2004) and others do, for a few reasons. First of all, the assumptions required to treat NHANES samples as validation data for NLSY cohorts are not credible given the mix of interview methods in the NLSY.[24] Secondly, rescaling BMI to adjust for systematic misreporting may change the BMI numbers at which we observe wage penalties, but it should not affect our ability to determine whether heavier workers are penalized at some point.

On a more fundamental level, it's not clear why even measured BMI should be considered the "true" variable of interest in studies of labor market outcomes. Employers rarely measure the body mass of their workers, but they easily observe whether workers appear heavier or lighter than is considered desirable. Therefore, body mass, measured or not, is likely just a proxy for the appearance of fatness (or thinness).[25]

2.2 Summary Statistics

Table 1 presents basic summary statistics for the white men and women in the sample. The appendix tables present additional summary statistics. The dependent variable in regressions that follow is the natural log of hourly wage.

[24] The critical assumption is that the distribution of actual measures conditional on reported is the same in both samples. Differences between interview modes suggest this assumption is violated even within NLSY cohorts. Furthermore, Han et al. (2009) suggest that NHANES respondents expect to be weighed, but NLSY respondents do not. See Courtemanche et al. (2015) for a more detailed discussion.

[25] Johansson et al. (2009) make a related point in their study of the effects of BMI, waist circumference and body fat on labor market outcomes in Finland. Rooth (2009) is noteworthy for using digitally manipulated photographs to estimate effects of appearing more or less fat.

Average log wage is around 2.3 for men and 2.2 for women, which translate to hourly wages of roughly 10 and 8.8.

The average respondent in the sample is roughly 24 years old.[26] In 2009, the average respondent (male or female) was 27 years old. Average years in the labor market is 4.5 for white men and four for women. The average male respondent in 2009 had been in the labor market for seven years, accumulating 6.4 years of actual work experience. The average woman in 2009 had been in the labor market for 6.2 years, accumulating 5.6 years of experience.

Average reported BMI for both genders exceeds 25, which is the threshold for being overweight. Less than 2% of white men and 4.2% of white women report being underweight (BMI $\leq$ 18.5). Over 57% of white men and nearly 42% of white women are overweight (BMI $\geq$ 25). 22% of white men and 20% of white women are obese (BMI $\geq$ 30). Almost 8% of men and over 10% of women report a BMI that qualifies as severely obese (BMI $\geq$ 35).

As seen in the appendix tables, movement between official BMI categories is not uncommon.[27] For either gender, moving to a heavier BMI category is more likely than moving to a lighter category. Additionally, the changes in BMI associated with changes in BMI categories (not shown) tend to be relatively large.[28] Therefore, it does not appear as though identification is

[26]Respondents were as young as 16 in period $t-2$, or 18 in t. This introduces the possibility that changes in BMI reflect changes in height for part of my sample and changes in weight for the rest; however, the results presented below are robust to excluding respondents under the age of 20.

[27]Roughly 9% of men and 8% of women move from one category to another between two consecutive years. If we consider three consecutive years, 15% of men and 13% of women in the sample have moved between categories at least once (not shown).

[28]The median percent change in BMI associated with a change in BMI category is 8.5

comes from small fluctuations in weight that employers would not notice.

3 Estimation

Recall that the differenced equations we're interested in take the form

$$\Delta w_{it} = \gamma \Delta w_{it-1} + \Delta X_{it}\beta + \Delta BMI_{it}\phi + \Delta BMI_{it-1}\phi_1 + \Delta\varepsilon_{it}. \qquad (3)$$

As noted above, the differenced GMM estimator uses second and further lagged levels as instruments for Δw_{it-1} and other endogenous variables. The GMM instruments enter as separate vectors for each year.[29] If a lag is missing for a given person and year, its value is set to zero. This allows the use of further lags when available without limiting observations unnecessarily.

Letting Z denote the matrix of all instruments and $\widehat{\epsilon}$ the vector of estimated residuals, the moment conditions are $E\left[Z'\widehat{\epsilon}\right] = 0$. These moments are estimated in Stata using XTABOND2.[30] All estimates use two-step efficient GMM, which produces robust standard errors, and apply the Windmeijer (2005) finite-sample variance correction.

percent for men, and 9.5 percent for women. Median changes in BMI are three to four times larger when categories change than they are otherwise.

[29]At $t = 3$, w_{i1} is an instrument for Δw_{i2}. At $t = 4$, w_{i2} and w_{i1} are instruments for Δw_{i3}, and so on.

[30]See Roodman (2006) for documentation.

3.1 Testing Assumptions

In Section 1, I assumed that the time-varying errors, ε_{it}, are not serially correlated. AB developed tests for this assumption. First-differenced regressions like equation (3) are $AR(1)$ by design. If there is serial correlation in ε, equation (3) will be at least $AR(2)$. I present tests for serial correlation with all of the results that follow.[31]

The results are also presented with tests of overidentifying restrictions. The Hansen J test examines the joint validity of all moment conditions. When BMI is treated as endogenous, difference-in-Hansen tests are used to evaluate the validity of wage lags and BMI lags separately.

The validity of lagged wage levels as instruments is independently evaluated by each of these tests. Lagged wage levels are valid instruments if ε_{it} are not serially correlated. If we fail to detect serial correlation that does exist, tests of overidentifying restrictions could still reject the validity of lagged wage instruments as long as some further lags are valid instruments.[32]

The validity of lagged BMI instruments are only evaluated directly by the overidentification tests; however, tests of serial correlation make the assumptions of overidentification tests more plausible. If BMI_{it} were correlated with ε_{it+1} but not ε_{it+2}, the second lags would not be valid instruments but the third lags would be. The validity of lagged BMI instruments would be

[31] Test results for $AR(3)$ and higher are available on request.

[32] The tests of serial correlation would have to miss a lot for lagged wage instruments to be untestable using overidentification. If equation (3) were $AR(2)$ but not $AR(3)$, w_{it-2} would not be a valid instrument for Δw_{it-1}, but w_{it-3} would be. If the regression were $AR(3)$ but not $AR(4)$, w_{it-4} would be a valid instrument, etc.

theoretically untestable only if BMI_{it} was correlated with ε_{it+1}, ε_{it+2}, etc.; but that seems unlikely unless those residuals are correlated with each other.

For example, BMI_{it} could be correlated with ε_{it+1} and later residuals due to health shocks that affect body mass more immediately than wages. But such health shocks would imply serial correlation in ε, unless the unobserved factors that were common to BMI_{it} and ε_{it+1} were independent of the factors common to BMI_{it} and ε_{it+2}. This strikes me as improbable; however, it is always possible that tests are too weak in practice to reject hypotheses that could be rejected in theory. Therefore, I consider the implications of health shocks for identification in Section 4.2.

3.2 Other Potential Problems with Instruments

The use of all lagged levels of wage and BMI variables produces a large number of instruments. A larger set of instruments improves efficiency, but not without a cost. As discussed by Roodman (2009) and others, using "too many" instruments overfits the endogenous variables, which biases coefficients towards OLS and weakens overidentification tests.

The results that follow restrict lags to the second through fifth. Using all available lags produces coefficients that are smaller in magnitude, which is consistent with adding weak instruments. Restricting lags further (e.g., excluding the fifth lag) has little effect on coefficient estimates.

3.3 Regression Specifications

All specifications presented in this paper model BMI_{it} as a vector of dummy variables for various levels of body mass. The use of dummy variables is motivated by the need for a simple specification that allows a non-linear relationship between BMI and wages.[33] Dummy variables are consistent with the idea that wage penalties are associated with weight exceeding levels that are considered desirable in the market. However, there is no theoretical reason to adopt one specification of BMI_{it} over any other.

When estimating regressions with dummy variables for BMI categories, the previous literature relied on categories defined by the WHO (overweight, obese, etc.); however, these categories were defined for the study of public health, not labor markets. As noted by Gregory and Ruhm (2011), wage penalties for high body mass may begin at points that fall between WHO cutoffs. Even if employers wanted to penalize workers based on the WHO categories, it's not clear how firms' imperfect assessments of body mass would line up with the imperfectly reported height and weight in the data.

The next section begins with specifications that use traditional BMI categories, but I also consider alternative BMI thresholds as a robustness check. Focusing on specifications of BMI_{it} that include one or two dummy variables, I estimate a large number of alternative models with thresholds ranging from 23 to 38. I then examine the robustness of coefficient estimates for a BMI

[33]In preliminary estimation, linear and polynomial specifications of BMI_{it} only produced statistically significant results in static models. Ignoring statistical significance, results using cubic polynomials are consistent with results from the preferred models.

category to other changes in the specification of BMI_{it}, as well as to the treatment of outliers in the distribution of wages. I also compare similar models using the specification tests of Bond et al. (2001) and Andrews and Lu (2001).[34] In the end, I find that specifications based on traditional cutoffs perform relatively well.

All of the dynamic models presented below include one lag of wage and one lag of BMI. None of the tests for serial correlation suggest that more lags are needed, and the tests of overidentification fail to suggest a problem with the instruments. Further lags of BMI are never statistically significant and do not change the basic results. The most obvious effect of adding a second lag of either BMI or the wage is a reduction of observations by 20-25%.

The suitability of a simple lag structure may be related to the youth of the sample. While one lag of wages or BMI may be insufficient in a sample of older workers, younger workers have less history in the market. It is also possible that the indirect effects of BMI_{it-2} and w_{it-2} on w_{it-1} make the addition of second lags redundant; however, adding further lags may simply ask too much of the data.

All regressions control for the local unemployment rate and incidence of obesity in the state, as well as dummy variables for region, urban residence, and being interviewed by telephone.[35] When lagged values of BMI are used,

[34]Both of these tests are based on comparisons of the Hansen J statistic. I use leave-one-out crossvalidation with these tests to ensure that the results are not peculiar to the estimation sample.

[35]Local unemployment and state identifiers are provided by the NLSY97 Geocode files. Percent obese in the state is estimated by the CDC.

the corresponding lag of the phone indicator is also included. Education enters as dummy variables for completing high school, some college, or college and beyond. I control for time using dummy variables for calendar year and years since labor market entry. No estimation in this paper is weighted.

I also control for actual experience in the labor market and it's square, as have Cawley (2004) and others.[36] To control for commitment to the labor market, I add interactions of experience with years since entry. Controlling for actual experience and its interactions makes the validity of lagged wages as instruments more plausible because lagged wages could reflect the accumulation of experience or commitment to the labor market.[37]

Accumulated experience could also affect BMI, as young workers who are steadily employed may have less time to invest in maintaining a healthy weight. Consistent with this possibility, I find that experience and BMI are positively correlated in my sample, even adjusting for age.[38] It's possible, therefore, that excluding experience would improperly attribute the effects of experience to body mass.[39]

[36]Controlling for experience is especially important to any study of the earnings of young women. As discussed by Altonji and Blank (1999), controlling for labor market experience has been critically important to understanding the gender wage gap. In addition to affecting estimates of the unexplained wage gap, controlling for experience is also necessary to avoid bias in other coefficients.

[37]Overidentification tests for lagged wage instruments improved in preliminary estimation with the use of actual experience and its interactions.

[38]For both men and women in my sample, experience has a positive, statistically significant correlation with BMI, overweight status, obesity and severely obesity. Controling for age, I find similar positive correlations, but the correlations between experience and severe obesity are no longer statistically significant. Changes in labor market experience are also positively and significantly correlated with changes in body mass.

[39]Lang and Manove (2011) present a similar (but more complete) argument in favor

On the other hand, labor market experience could be endogenous in any wage equation. Experience could also be affected by unobserved characteristics that affect body mass, or affected more directly by high body mass. Fortunately, the use of a well-established instrument, potential experience (age−schooling−six), for actual experience should alleviate concern about bias either case.[40]

Finally, there are a number of potential confounders, such as occupation or hours worked, that are excluded from the preferred models due to their likely endogeneity. I find that the main results are robust to the addition of most potentially endogenous confounders.[41] The estimated effects of BMI on the wages of women do fall when controls for occupation are included; however, the exogeneity of both current and lagged occupation can be rejected. Even second lags for some occupations can be rejected.

---

of controlling for education when analyzing the black/white wage gap. They show that black men acquire more education than white men of the same measured ability. As a result, work that excludes education on the grounds that it is endogenous and affected by anticipated discrimination understate the racial wage gap.

[40]I use the square of potential experience and its interactions where appropriate. Using potential experience as a traditional instrument in dynamic panel data models means using it's first difference, and most of the variance in this difference is due to exogenous variation in interview dates.

[41]The results are robust to controls for marital status, the number and ages of children, hours worked, tenure, on-the-job training, and employer-provided health insurance. Angrist and Krueger (1999), among others, note that a lack of robustness to changes in the included variables could suggest sensitivity unobservable characteristics.

4 Results

Section 4.1 presents the main results, as well as initial tests of identifying assumptions; and then compares the preferred dynamic models to simpler regressions. Section 4.2 presents additional tests of the identification and robustness of the main results.

4.1 The Preferred Models

As discussed above, BMI_{it} is modeled using dummy variables for exceeding various BMI thresholds. For both men and women, I present specifications based on the familiar categories of overweight, obese and severely obese before discussing alternative specifications of BMI_{it}. I then compare the preferred dynamic models to simpler dynamic and static models.

4.1.1 Results for White Men

Table 2 presents results for white men from various models that use dummy variables indicating overweight, obesity or severe obesity.[42] First of all, note that the tests presented in the lower panel are all consistent with the identifying assumptions discussed above. All of the equations are AR(1) due to first-differencing; but none of them are AR(2), which they would be if there

[42]All specifications in Tables 2 and 3 use lagged indicators of overweight, obesity and severe obesity as instruments to facilitate comparisons. I also use lags of three BMI dummies as instruments when alternative BMI cutoffs are considered, with each alternative category replacing the closest WHO category.

were serial correlation in the residuals.[43] Furthermore, none of the tests of overidentifying restrictions reject the validity of the GMM instruments. The p-values for Hansen and difference-in-Hansen tests are all well above the conservative threshold of 0.25 suggested by Roodman (2009).

The coefficient estimates in Table 2 suggest that white men face a penalty of roughly 17% for having been severely obese in the previous year. The coefficients on lagged severe obesity range from -0.165 (0.060) in column 4 to -0.172 (0.059) in column 5. No other BMI variable has a statistically significant coefficient in this table, and there is no evidence of current BMI having any effect on the wages of white men.[44]

The results for men also support the inclusion of an autoregressive term in the wage equations, and I find that controlling for lagged wages is essential for identification. In specifications that include severe obesity, coefficients on lagged log wages are between 0.072 (0.041) and 0.077 (0.042). Specifications that exclude lagged wages but are otherwise similar to those in Table 2 (not shown) are at least AR(2). Consistent with Cawley's (2004) point that lagged BMI variables are not likely to be valid instruments in the presence of serial correlation, the second lags of BMI are rejected by overidentification tests in static wage equations.

When I consider alternative dummy variable specifications of BMI_{it} for white men, I find that the penalty for lagged severe obesity is remarkably

[43] I also found no evidence of the residuals being $AR(3)$ or higher.

[44] Results for the model that specified BMI_{it} using overweight and obese are excluded. They are consistent with those in columns 2 and 3.

robust.[45] There is no robust evidence of an effect at lower levels of current or lagged body mass for white men. Furthermore, the specification test proposed by Andrews and Lu (2001) supports models that use severe obesity over similar models that use nearby cutoffs.[46]

I select a preferred specification from Table 2 using the test of parameter restrictions proposed by Bond et al. (2001).[47] The only restricted specification that cannot be rejected in favor of the full specification (column 1) is the specification in column 5, which includes dummy variables for being overweight and severely obese.[48] Since the specification in column 5 is simpler than that in column 1, it is my preferred specification in what follows.

4.1.2 Results for White Women

The results presented in Table 3 show that white women face a penalty for a lagged BMI in (or above) the overweight category. The coefficient on lagged overweight status is -0.082 (0.040) in column 2, and -0.093 (0.044) in column 6. The Andrews-Lu test suggest that the specification in columns 2 and 6 are preferred to others that model BMI_{it} using the same number of traditional BMI categories. The specification in column (6), which models BMI_{it} using dummy variables for reported overweight and severe obesity is also the same

[45] The coefficient on lagged severe obesity is qualitatively similar and statistically significant in all models with a lower cutoff below 33. Coefficients on lagged BMI$\geq$34.5 are similar, but smaller and statistically significant in fewer models.

[46] When comparing models with the same number of variables and instruments, the test of Andrews and Lu (2001) selects the model that minimizes the Hansen J statistic.

[47] Bond et al. (2001) show that differences between the J statistics of restricted and unrestricted models are χ^2 with degrees of freedom equal to the number of restrictions.

[48] Columns 2, 3 and 6 are rejected at a 5% level. Column 4, at a 10% level.

as the preferred specification for white men.

Alternative BMI thresholds are more relevant for white women than for white men. The penalty for a lagged BMI ≥ 24.5 is robust across all specifications with a second cutoff at or above 27, or without a second BMI variable. As seen in Table 4, the coefficients on lagged BMI ≥ 24.5 (columns 3-5) are similar to, but larger than analogous coefficients on lagged overweight status (columns 1 and 2).[49]

When I consider alternative thresholds in the heavier part of the BMI distribution, I find evidence of additional penalties for a BMI ≥ 37. This suggests penalties for being in the heaviest eight percent of the sample, as do the results for severely obese white men.[50] The results in columns 2 and 5 of Table 4 suggest a penalty for a current BMI ≥ 37 of 13-14%. Furthermore, the coefficient on a lagged BMI ≥ 37 is -0.099 (0.038) in column 5.[51]

The Hansen J statistics in Table 4 support the use of reported overweight status instead of the lower threshold of 24.5; however, the direction of this difference is not robust to the treatment of outliers or to minor changes in the set of instruments used. In either case, the results suggest that women who reported a BMI at or near the threshold for overweight in the previous year face a penalty of 9-12%.

Regardless of specification, I find that lagged wages have larger effects

[49] Coefficients on lagged overweight status are statistically significant in roughly half as many models as coefficients on lagged BMI ≥ 24.5.

[50] 7.9% of women in the sample report a BMI ≥ 37. 7.8% of men report a BMI ≥ 35.

[51] The number of instruments changes when a higher threshold of 37 is used instead of the standard cutoff of 35, making comparisons of Hansen J statistics difficult.

on current wages for white women than for white men. The coefficients on w_{it-1} range from 0.211 (0.057) to 0.231 (0.057) for women, compared to 0.077 (0.042) for men. As in Table 2, there is no evidence of serial correlation in the residuals of any model shown in Table 3 or 4; however, models that exclude lagged wages (not shown) are at least AR(2). Furthermore, none of the overidentification tests in Tables 2 through 4 cast doubt on the validity of the instruments.

On a more meaningful level, autoregressive wage equations imply that further lags of body mass (BMI_{it-2}, BMI_{it-3}, etc.) affect wages even when only BMI_{it-1} enters the model directly. The lagged wage, w_{it-1}, is a function of BMI_{it-2} and w_{it-2}, which is affected by BMI_{it-3} and w_{it-3}, and so on. This is an important advantage of dynamic models over the static models of previous work: Wages are allowed to respond slowly to changes in BMI, resulting in penalties that persist and accumulate over time.

As an example, consider two women who enter the market at $t-2$. One has a BMI of 38 and the other has a BMI of 23. The heavier woman loses weight and the other gains weight so that both are overweight (but not obese) in years $t-1$ and t. In $t-1$, the woman who had been heavier faces a 20% penalty according to column (5) of Table 4, while the woman who had been lighter faces no statistically significant penalty for her recent weight gain. In year t, both women have been overweight (but not obese) for two years; however, one faces a penalty of 10% and the other faces a penalty of up to

15%.[52] Even using the more conservative specification of column (1), the woman who was heavier at $t-2$ would face a penalty of 11.4% in year t, while the other woman faces a penalty of 9.3%.

The negative effects of high body mass also accumulate over time. A woman entering the market with a BMI over 37 faces an initial penalty of 13%. If she does not lose weight, she will again face the 13% penalty for current body mass in her second year, plus a 20% penalty for her past body mass. In her third year with a BMI over 37, she will be penalized an additional 4.5% for her BMI two years ago, on top of the 33% penalty for her BMI in t and $t-1$. A woman with a reported BMI near the threshold for being overweight faces a penalty of 9-12% after one year that approaches 12-16% over time.

4.1.3 Comparisons to Simpler Models

Tables 5A and 5B compare the preferred dynamic models to OLS regressions and a dynamic model that assumes ΔBMI_{it} is exogenous. The OLS results for men in Table 5A suggest that overweight men are paid more than lighter peers while severely obese men are paid less. The OLS results in Table 5B suggest that being overweight (upper panel) or having a BMI ≥ 24.5 (lower panel) is associated with lower wages for white women; however, this negative association is due to lagged BMI, not current.

[52] The estimated effect of $BMI_{it-2} \geq 37$ in t is -0.044 (0.020), the effect in $t-1$ multiplied by the coefficient on Δw_{it-1}. If the woman was also overweight in $t-1$, her combined penalty in t is -0.148 (0.056).

The autoregressions presented in the second columns of Tables 5A and 5B remove any bias due to individual fixed effects or the omission of lagged wages, but treat ΔBMI_{it} and ΔBMI_{it-1} as exogenous. Tests of overidentifying restrictions (not shown) reject this assumption, suggesting that changes in BMI are correlated with time-varying unobservables. The fact that the negative effects of body mass found in the preferred specifications are not found in the second column of either table is consistent with time-varying unobservables, such as time spent sitting at a desk, that are positively correlated with both wages and body mass.

Finally, note that the error terms in the OLS regressions are serially correlated, while the residuals in the dynamic models are not. The serial correlation of residuals in the static models supports the argument of Cawley (2004) against the use of lagged BMI variables as instruments in static wage regressions. In contrast, the dynamic wage equations explicitly model the correlation between past and current wages that creates the serial correlation in static wage regressions. As a result, we see no evidence of *residual* serial correlation in the dynamic models.

4.2 Robustness Tests and Potential Sources of Bias

The identifying assumptions in this paper have already received considerable scrutiny from tests of serial correlation and overidentifying restrictions. The results present above are also robust to changes in the treatment of outlier, as well as minor changes in model specification. This section presents additional tests of identifying assumptions, and considers the potential effects of sample selection on the results.

4.2.1 Health Shocks & Identification

There are two justifications for further examination of the identifying assumptions in this paper. First, as is always the case, it is possible that a test presented in Section 4.1 failed to reject an assumption that should have been rejected.[53] Secondly, as discussed in Section 3.1, the validity of lagged levels of BMI as instruments would be theoretically untestable if BMI_{it} were somehow correlated with all future wage residuals (ε_{it+1}, ε_{it+2}, etc.) without those residuals being correlated with each other.

The most obvious reason why BMI_{it} may be correlated with future wage residuals is that random health shocks could affect body mass immediately, but have delayed effects on wages that are independent of BMI_{it}.[54] Although health shocks that cause BMI_{it} to be correlated with both ε_{it+1} and ε_{it+2} should also cause serial correlation in ε, which I've tested for, I examine the possibility that my main results are biased by such shocks in Table 6.

Table 6 compares the preferred models from Section 4.1 to models that add indicators of current and lagged general health. The self-reported measure of health that is available in the data is not ideal; however, it is corre-

[53]Fortunately, tests of serial correlation and overidentifying restrictions do reject hypotheses that we expect to be rejected, but type II errors are still possible.

[54]I also examined the possibility that hard-to-observe individual ability affects both BMI_{it} and future wage growth due to employer learning or human capital accumulation. My analysis, which I ommit for brevity's sake, found no evidence of bias related to time-varying effects of ability.

lated with body mass as one would expect.[55] If the main results are biased by health shocks that are correlated with changes in reported health, those results should not be robust to the inclusion of changes in health. The results would also not be robust to the addition of health variables if health were an intermediate outcome through which body mass affected wages. Therefore, the falsification test in Table 6 could suggest a problem with my identifying assumptions even when none exists.

The preferred models for men and women are quite robust to the inclusion of changes in general health status. Regardless of whether I add only the most recent changes or also include lagged changes in health, the coefficients on BMI are similar.[56] Furthermore, none of the overidentification tests change between columns in a way that suggests health shocks affect the identification of the BMI variables, or the overall identification of the model.

4.2.2 Sample Selection

The previous literature provides some evidence of selection into employment varying with body mass.[57] Since the estimation in this paper requires three consecutive years of wages, some investigation of sample selection bias is war-

[55] Respondents were asked to evaluate their health on a five-point scale from "excellent" (1) to "poor" (5). Higher levels of BMI are negatively correlated with "excellent" or "very good" health, and positively correlated with "fair" or "poor" health.

[56] I obtain similar results from models with up to five lags of the health indicators, and models that use residuals of lagged BMI on health as instruments. Results for BMI are also not sensitive to assumptions about the exogeneity of health.

[57] Rooth (2009) finds that body mass affects call-back rates for both men and women in Sweden, while Caliendo and Lee (2013) find negative effects on employment only for obese women in Germany.

ranted. That said, a differenced estimator would only be biased by selection on time-varying unobservables. Selection on unobservables that are constant over time is addressed by differencing.

Following Semykina and Wooldridge (2010), I estimate probits for the probability of being in the sample in year t with valid observations for t, $t-1$ and $t-2$.[58] I then add the inverse Mills ratios and their interactions with time dummies to the wage regressions. The hypothesis of no selection is rejected if coefficients on the Mills ratios and their interactions are jointly significant.

These tests do not suggest bias from sample selection. None of the coefficients on the inverse Mills ratios or their interactions are statistically significant. The p-values on tests of joint significance are 0.26 for men and 0.29 for women.

4.2.3 Further Examination of the Strength of Instruments

Although preliminary regressions support the assumption that BMI_{it-2} and other recent lags predict ΔBMI_{it}, one might still worry that there is not enough variation in the lagged levels to identify coefficients on both ΔBMI_{it} and ΔBMI_{it-1}. In that case, the coefficients in my preferred models, especially for ΔBMI_{it}, would be biased toward zero.[59]

One way to address a potential lack of variation would be adding lags

[58] I include AFQT, which is differenced out of wage equations, in the probit estimates.

[59] More precisely, the coefficients would be biased toward those in the second columns of Tables 5A and 5B, which are small and statistically insignificant.

of continuous BMI variables to the instrument matrix. This would exploit variation in past BMI that is not captured by dummy variables. When I add lags of BMI to the set of instruments, however, the results (not shown) are similar to those presented above.[60]

I also compared the preferred specifications to restricted specifications that exclude either ΔBMI_{it} or ΔBMI_{it-1}.[61] If there is an effect of ΔBMI_{it} that we haven't observed due to insufficient variation in the instruments, the coefficients on ΔBMI_{it} should change when ΔBMI_{it-1} is excluded.[62]

The results (not shown) do not suggest that the preferred models are biased by weak instruments. Dropping ΔBMI_{it-1} from wage equations does not reveal previously unidentified effects of ΔBMI_{it}. In fact, the estimated penalty white women face for a current BMI ≥ 37 is only statistically significant when lagged BMI is included in the regression. Furthermore, the specification test of Bond et al. (2001) rejects the restricted models for both men and women in favor of models that include both ΔBMI_{it} and ΔBMI_{it-1}.

[60] I add only the second and third lags of continuous BMI to limit the increase in the number of instruments.

[61] Kropfhäußer and Sunder (2015), citing an early version of the current paper, estimate models that only include BMI variables from one year. They use German data that includes wages every two years and BMI in off years. With t indexing years, their model takes the form: $w_{it} = \gamma w_{it-2} + X_{it}\beta + BMI_{it-1}\phi + \nu_{it}$, where BMI_{it-1} is quadratic.

[62] The coefficients could also change because ΔBMI_{it} is correlated with ΔBMI_{it-1}.

5 Conclusions

The results of this paper demonstrate the importance of using dynamic models when considering effects of body mass on wages. I find that past body mass affects the wages of young workers more often than current body mass. Furthermore, current wages are affected by lagged wages, which are affected by further lags of body mass and wages. Therefore, my results suggest that the penalties for a high body mass persist and even accumulate over time, especially for women.

I find that white men are penalized for past severe obesity, while white women begin to face penalties slightly before the threshold for being overweight. Women are penalized for a past BMI ≥ 24.5, with additional penalties for past or current body mass that begin above the threshold for severe obesity. Considering the distribution of BMI in the sample, the results suggest that over 40 percent of young white women face some penalty for their weight, while roughly eight percent of men face penalties.

Consistent with the youth of the sample, the penalties for high body mass identified in this paper do not appear to be driven by changes in health or selection into the labor market. The results are robust to the addition of controls for general health, and tests for sample selection bias provide no cause for concern. Furthermore, the results survive a number of other robustness tests, as well as tests of identifying assumptions.

While this paper cannot identify the reasons workers are penalized for high body mass, the estimation is more consistent with theories of discrimination that incorporate labor market frictions than is most previous work. For example, models in which discrimination affects labor market search, like those in Bowlus and Eckstein (2002) and Lang et al. (2005), would suggest penalties for past body mass. Wages would be affected by body mass when workers found their current jobs, and lower wages on one job would lead to lower reservation wages as workers look for better jobs.

Finally, persistent effects of high body mass on wages are also consistent with effects on occupational selection. The models presented in this paper exclude occupation due to its endogeneity, and tests reject the exogeneity of even lagged occupation variables. These test results suggest that the occupational selection observed in the NLSY79 by Lakdawalla and Philipson (2007) and Harris (2015) may still affect the wages of young workers who entered the labor market in more recent decades.

References

Altonji, J. G., P. Bharadwaj, and F. Lange (2012). Changes in the characteristics of american youth: Implications for adult outcomes. *Journal of Labor Economics 30*(4), 783–828.

Altonji, J. G. and R. M. Blank (1999). Race and gender in the labor market. *Handbook of labor economics 3*, 3143–3259.

Altonji, J. G. and C. R. Pierret (2001). Employer learning and statistical discrimination. *Quarterly Journal of Economics 116*(1), 313–350.

Andrews, D. W. and B. Lu (2001). Consistent model and moment selection procedures for gmm estimation with application to dynamic panel data models. *Journal of Econometrics 101*(1), 123–164.

Arellano, M. (2003). *Panel Data Econometrics*. New York: Oxford University Press.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies 58*(2), 277–297.

Arellano, M. and B. Honoré (2001). Panel data models: some recent developments. *Handbook of econometrics 5*, 3229–3296.

Averett, S. and S. Korenmann (1996). The economic reality of the beauty myth. *The Journal of Human Resources 31*, 304–330.

Baum, C. L. and W. F. Ford (2004). The wage effects of obesity: A longitudinal study. *Health Economics 13*(9), 885–899.

Behrman, J. and M. Rosenzweig (2001). The returns to increasing body weight.

Bjerk, D. (2008). Glass ceilings or sticky floors? statistical discrimination in a dynamic model of hiring and promotion*. *The Economic Journal 118*(530), 961–982.

Blau, F. D. and L. M. Kahn (2013). The feasibility and importance of adding measures of actual experience to cross-sectional data collection. *Journal of Labor Economics 31*(S1), S17–S58.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics 87*(1), 115–143.

Bond, S. R. (2002). Dynamic panel data models: A guide to micro data methods and practice. *Portuguese Economic Journal 1*(1), 141–162.

Bond, S. R., C. Bowsher, and F. Windmeijer (2001). Criterion-based inference for gmm in autoregressive panel data models. *Economics Letters 28*(3).

Bowlus, A. J. and Z. Eckstein (2002). Discrimination and skill differences in an equilibrium search model. *International Economic Review 43*(4), 1309–1345.

Cawley, J. (2004). The impact of obesity on wages. *Journal of Human Resources 39*(2), 451–474.

Cawley, J., E. Han, and E. C. Norton (2011). The validity of genes related to neurotransmitters as instrumental variables. *Health Economics 20*(8), 884–888.

Chen, A. J. (2012). When does weight matter most? *Journal of Health Economics 31*(1), 285–295.

Courtemanche, C., J. C. Pinkston, and J. Stewart (2015). Adjusting body mass for measurement error with invalid validation data. *Economics & Human Biology 19*, 275–293.

Goldman, D., D. Lakdawalla, and Y. Zheng (2010). Food prices and the dynamics of body weight. Working Paper 15096, NBER.

Gregory, C. A. and C. J. Ruhm (2011). Where does the wage penalty bite? In *Economic aspects of obesity*, pp. 315–347. University of Chicago Press.

Hamermesh, D. S. (2011). *Beauty pays: Why attractive people are more successful*. Princeton University Press.

Han, E., E. C. Norton, and L. M. Powell. (2011). Direct and indirect effects of teenage body weight on adult wages. *Economics and Human Biology 9*(4), 381–392.

Han, E., E. C. Norton, and S. C. Stearns (2009). Weight and wages: fat versus lean paychecks. *Health Economics 18*(5), 535–548.

Harris, M. (2015). The impact of body weight on occupational mobility and career development.

Holtz-Eakin, D., W. Newey, and H. S. Rosen (1988). Estimating vector autoregressions with panel data. *Econometrica 56*, 1371–1395.

Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (1993). Earnings losses of displaced workers. *The American Economic Review*, 685–709.

Johansson, E., P. Böckerman, U. Kiiskinen, and M. Heliövaara (2009). Obesity and labour market success in finland: The difference between having a high bmi and being fat. *Economics & Human Biology 7*(1), 36–45.

Kahn, L. B. (2010). The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics 17*(2), 303–316.

Kline, B. and J. L. Tobias (2008). The wages of bmi: Bayesian analysis of a skewed treatment-response model with non-parametric endogeneity. *Journal of Applied Econometrics 23*(6), 767–793.

Kropfhäußer, F. and M. Sunder (2015). A weighty issue revisited: the dynamic effect of body weight on earnings and satisfaction in germany. *Applied Economics 47*(41), 4364–4376.

Lakdawalla, D. and T. Philipson (2007). Labor supply and body weight. *Journal of Human Resources 42*, 85–116.

Lang, K. and M. Manove (2011). Education and labor market discrimination. *The American Economic Review 101*(4), 1467–1496.

Lang, K., M. Manove, and W. T. Dickens (2005). Racial discrimination in labor markets with posted wage offers. *The American economic review 95*(4), 1327–1340.

Lee, I.-M., L. Djoussé, H. D. Sesso, L. Wang, and J. E. Buring (2010). Physical activity and weight gain prevention. *JAMA 303*(12), 1173–1179.

Lewis, C. E., D. R. Jacobs, H. McCreath, C. I. Kiefe, P. J. Schreiner, D. E. Smith, and O. D. Williams (2000). Weight gain continues in the 1990s: 10-year trends in weight and overweight from the cardia study. *American Journal of Epidemiology 151*(12), 1172–1181.

Michaud, P.-C. and A. V. Soest. (2008). Health and wealth of elderly couples: Causality tests using dynamic panel data models. *Journal of Health Economics 27*(5), 1312–1325.

Neal, D. A. (1999). The complexity of job mobility among young men. *Journal of Labor Economics 17*(2), 237–61.

Ng, S. W., E. C. Norton, D. K. Guilkey, and B. M. Popkin (2010). Estimation of a dynamic model of weight. Technical report, NBER Working Papers 15864.

Ogden, C. L. and M. D. Carroll (2010). Prevalence of overweight, obesity, and extreme obesity among adults: United states, trends 1960–1962 through 2007–2008. *National Center for Health Statistics 6*, 1–6.

Oreopoulos, P., T. von Wachter, and A. Heisz (2012). The short-and long-term career effects of graduating in a recession. *American Economic Journal: Applied Economics 4*(1), 1–29.

Pischke, J.-S. (1995). Measurement error and earnings dynamics: Some estimates from the psid validation study. *Journal of Business & Economic Statistics 13*(3), 305–314.

Roodman, D. M. (2006). How to do xtabond2: An introduction to 'difference' and 'system' gmm in stata. working paper 103, Center for Global Development.

Roodman, D. M. (2009). A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics 71*(1), 135–158.

Rooth, D.-O. (2009). Obesity, attractiveness, and differential treatment in hiring: A field experiment. *Journal of Human Resources 44*(3), 710–735.

Sabia, J. J. and D. I. Rees (2012). Body weight and wages: Evidence from add health. *Economics & Human Biology 10*(1), 14–19.

Semykina, A. and J. M. Wooldridge (2010). Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics 157*(2), 375–380.

Topel, R. H. and M. P. Ward (1992). Job mobility and the careers of young men. *The Quarterly Journal of Economics 107*(2), 439–479.

Williamson, D. F., H. S. Kahn, P. L. Remington, and R. F. Anda (1990). The 10-year incidence of overweight and major weight gain in us adults. *Archives of Internal Medicine 150*(3), 665.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step gmm estimators. *Journal of Econometrics 126*(1), 25–51.

A Data

This appendix describes the selection of the estimation sample, and presents more detailed summary statistics.

The sample was first restricted to white respondents in the NLSY97. This leaves 2,702 white men and 2,530 white women. Respondents who reported being in the military were then dropped, leaving 34,522 person/year observations for 2,470 men, and 31,378 observations for 2,481 women.[63]

Restricting observations to those in which the respondent has entered the labor market reduces the number of observations to 12,368 for 1,771 white men, and 10,008 for 1,566 women. Only the primary (current or most recent) job is used from each interview. Observations in which a woman reported being pregnant in the current year (since the last interview) or previous year were dropped, reducing observations to 8,600 for 1,558 women. Finally, 5 observations for men and 2 for a woman were dropped because the absolute value of the change in log wages was greater than 6.5.[64]

The preferred dynamic specifications require three consecutive observations with non-missing values of wage and BMI. Limiting observations to those that could be from one of three consecutive years leaves 9,037 observations for 1,473 men and 5,408 observations for 1,060 women.

Appendix tables A1 and A2 present summary statistics for men and women that are not presented in Table 1. As expected, the sample is largely

[63] This includes 63 observations in which a respondent reported a military occupation despite not being otherwise identified as in the military.

[64] These observations were obvious outliers in the distribution of log wage changes. One of the wage observations in each case was below $0.2. Otherwise, observations that might appear to be outliers in the distribution of wages were not dropped from the sample.

urban. The differences in urbanicity between men and women, as well as most of the differences in education, appear to be due to how men and women enter the labor market. Looking at the entire sample (not shown), instead of focusing on those in the estimation sample, reveals no difference by gender in urbanicity and much smaller differences in education.

Table 1. Summary Statistics

	Mean	Std. Dev.	Min	Max
White Men				
Wage	19.887	356.121	0.039	23,883.93
Log Wage	2.332	0.619	-3.252	10.081
BMI	26.721	5.524	12.838	63.313
Underweight	0.018	0.134	0	1
Overweight	0.573	0.495	0	1
Obese	0.222	0.415	0	1
Severely Obese	0.078	0.269	0	1
Age	23.711	2.693	16	30
Phone Interview	0.108	0.310	0	1
Yrs since LM Entry	4.474	2.658	1	14
Yrs in 2009	7.012	2.563	3	14
Actual Experience	4.172	2.459	0.75	13.058
Exp in 2009	6.398	2.413	1.846	13.058
White Women				
Wage	11.051	21.528	0.046	774.08
Log Wage	2.196	0.563	-3.069	6.652
BMI	25.810	6.665	10.962	72.620
Underweight	0.042	0.201	0	1
Overweight	0.417	0.493	0	1
Obese	0.202	0.401	0	1
Severely Obese	0.103	0.304	0	1
Age	23.963	2.554	16	30
Phone Interview	0.103	0.304	0	1
Yrs since LM Entry	4.053	2.491	1	13
Yrs in 2009	6.187	2.510	3	13
Actual Experience	3.793	2.293	0.75	12.769
Exp in 2009	5.625	2.334	1.558	12.769

The sample for this table includes all observations that are used as t , t- 1, or t -2 in the main estimation. There are 9,037 observations for 1,473 white men; and 5,408 observations for 1060 white women.

Table 2. Effects of Past and Current BMI on the Log Wages of White Men

Results for Traditional BMI Categories

	(1)	(2)	(3)	(4)	(5)	(6)
L.ln(wage)	0.0722*	0.0605	0.0648	0.0769*	0.0720*	0.0768*
	(0.0414)	(0.0410)	(0.0416)	(0.0420)	(0.0412)	(0.0421)
Overweight	-0.1187	-0.0977	…	…	-0.1190	…
	(0.1198)	(0.1224)			(0.1207)	
L.Overweight	0.0540	0.0477	…	…	0.0549	…
	(0.0491)	(0.0456)			(0.0470)	
Obese	0.0112	…	0.0054	…	…	-0.0051
	(0.0979)		(0.1021)			(0.1012)
L.Obese	0.0030	…	0.0175	…	…	0.0061
	(0.0567)		(0.0576)			(0.0600)
Severely Obese	0.0465	…	…	0.0745	0.0557	0.0360
	(0.0927)			(0.0938)	(0.0925)	(0.0951)
L.Severely Obese	-0.1699***	…	…	-0.1649***	-0.1722***	-0.1668**
	(0.0632)			(0.0598)	(0.0594)	(0.0680)
AR(1): *z*-statistic	-5.008	-4.968	-4.995	-5.003	-4.998	-4.994
p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
AR(2): *z*-statistic	-0.542	-0.685	-0.693	-0.595	-0.568	-0.603
p-value	0.588	0.494	0.489	0.552	0.570	0.546
Hansen *J* Statistic	128.5	144.4	146.3	136.7	130.3	135.3
Hansen test *p*-value	0.802	0.546	0.501	0.718	0.803	0.707
Diff-in-Hansen Tests for Exogeneity of Subsets of GMM Instruments (*p*-values)						
ln(wage) lags	0.377	0.367	0.395	0.341	0.347	0.339
BMI cat. Lags	0.770	0.516	0.533	0.662	0.738	0.678

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 . There are 5,897 observations for 1,473 men. All specifications have a total of 196 instruments, with lags of all three BMI dummies used as instruments in each case. Regressions also control for the local unemployment rate and percent obese in the state, as well as dummies for region, urban residence, being interviewed over the phone and its lag, completing HS, some college, and college or beyond, calendar year and years since labor market entry. To control for commitment to the labor market I include actual experience and its interactions with years since entry; however, I treat experience as endogenous, using potential experience and its interactions as instruments.

Table 3. Effects of Past and Current BMI on the Log Wages of White Women

Results for Traditional BMI Categories

	(1)	(2)	(3)	(4)	(5)	(6)
L.ln(wage)	0.2110***	0.2308***	0.2158***	0.2209***	0.2237***	0.2223***
	(0.0567)	(0.0568)	(0.0542)	(0.0557)	(0.0577)	(0.0575)
Overweight	0.0420	0.0760	…	…	0.0509	0.0658
	(0.1007)	(0.0547)			(0.0957)	(0.0549)
L.Overweight	-0.0781	-0.0819**	…	…	-0.0625	-0.0931**
	(0.0519)	(0.0402)			(0.0466)	(0.0436)
Obese	0.0712	…	0.0937	…	0.0785	…
	(0.0840)		(0.0649)		(0.0814)	
L.Obese	-0.0319	…	-0.0032	…	-0.0085	…
	(0.0582)		(0.0477)		(0.0563)	
Severely Obese	0.0166	…	…	0.0267	…	0.0155
	(0.0589)			(0.0672)		(0.0612)
L.Severely Obese	-0.0640	…	…	-0.0579	…	-0.0503
	(0.0718)			(0.0647)		(0.0661)
AR(1): *z*-statistic	-4.863	-4.994	-4.937	-4.928	-4.941	-4.907
p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
AR(2): *z*-statistic	-0.405	-0.162	-0.329	-0.455	-0.261	-0.405
p-value	0.686	0.871	0.742	0.649	0.794	0.686
Hansen *J* Statistic	111.8	114.6	115.2	121.2	113.4	112.2
Hansen test *p*-value	0.936	0.943	0.938	0.872	0.938	0.947
Diff-in-Hansen Tests for Exogeneity of Subsets of GMM Instruments (*p*-value)						
ln(wage) lags	0.704	0.667	0.745	0.559	0.775	0.740
BMI cat. Lags	0.931	0.939	0.912	0.855	0.928	0.958

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 . There are 3,154 observations for 1,060 women. All specifications have a total of 187 instruments, with lags of all three BMI dummies used as instruments in each case. Regressions also control for the local unemployment rate and percent obese in the state, as well as dummies for region, urban residence, being interviewed over the phone and its lag, completing HS, some college, and college or beyond, calendar year and years since labor market entry. To control for commitment to the labor market I include actual experience and its interactions with years since entry; however, I treat experience as endogenous, using potential experience and its interactions as instruments.

Table 4. Effects of Past and Current BMI on the Log Wages of White Women

Results for Alternative BMI Categories

	Lower BMI Variable: Overweight		Lower BMI Variable: BMI ≥ 24.5		
	(1)	(2)	(3)	(4)	(5)
L.ln(wage)	0.2223***	0.224***	0.214***	0.216***	0.220***
	(0.0575)	(0.053)	(0.059)	(0.059)	(0.053)
Lower BMI Var.	0.0658	0.084	0.050	-0.003	0.018
	(0.0549)	(0.071)	(0.148)	(0.145)	(0.085)
L.(Lower BMI Var.)	-0.0931**	-0.087**	-0.112***	-0.123***	-0.104**
	(0.0436)	(0.041)	(0.039)	(0.041)	(0.043)
Severely Obese	0.0155	…	…	0.008	…
	(0.0612)			(0.060)	
L.Severely Obese	-0.0503	…	…	-0.080	…
	(0.0661)			(0.065)	
BMI ≥ 37	…	-0.138**	…	…	-0.131**
		(0.058)			(0.056)
L.(BMI ≥ 37)	…	-0.069	…	…	-0.099*
		(0.057)			(0.058)
AR(1): *z*-statistic	-4.907	-5.06	-4.91	-4.91	-5.00
p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
AR(2): *z*-statistic	-0.405	-0.47	-0.48	-0.63	-0.62
p-value	0.686	0.637	0.634	0.528	0.534
Hansen *J* statistic	112.2	101.27	120.3	117.9	105.3
Hansen test *p*-value	0.947	0.989	0.885	0.891	0.976
Diff-in-Hansen Tests for Exogeneity of Subsets of GMM Instruments (*p*-values)					
ln(wage) lags	0.740	0.749	0.590	0.53	0.472
BMI cat. Lags	0.958	0.993	0.856	0.888	0.983

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 . There are 3,154 observations for 1,060 women. All specifications use lags of three BMI dummy variables as GMM instruments. (See text for further detail.) There are 187 instruments in columns (1), (3) and (4); and 185 in (2) and (5) where BMI ≥ 37 is used. See Table 3 for a description of control variables.

Table 5A. Preferred Specification Compared to Simpler Models

White Men

	Preferred Dynamic Specification	Dynamic w/ BMI exogenous	OLS W/out Lags	OLS W/ Lags
L.ln(wage)	0.0720*	0.1213***	…	…
	(0.0412)	(0.0470)		
Overweight	-0.1190	0.0305	0.0600***	0.0332
	(0.1207)	(0.0317)	(0.0165)	(0.0266)
L.overwt	0.0549	0.0115	…	0.0369
	(0.0470)	(0.0312)		(0.0262)
Severely Obese	0.0557	0.0314	-0.0646**	-0.0116
	(0.0925)	(0.0382)	(0.0262)	(0.0493)
L.(Severely Obese)	-0.1722***	-0.0525	…	-0.0799
	(0.0594)	(0.0609)		(0.0515)
AR(1): *z*-statistic	-4.998	-4.746	10.34	10.35
p-value	< 0.001	< 0.001	< 0.001	< 0.001
AR(2): *z*-statistic	-0.568	-0.523	11.71	11.65
p-value	0.570	0.601	< 0.001	< 0.001
Number of Instruments	196	88	…	…
Hansen test (*p*-value)	0.803	0.766	…	…

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 . There are 5,897 observations for 1,473 men. Specifications are as described previously, except for BMI variables being treated as exogenous in the later three columns.

Table 5B. Preferred Specification Compared to Simpler Models

White Women

	Preferred Dynamic Specification	Dynamic w/ BMI exogenous	OLS W/out Lags	OLS W/ Lags
WHO BMI Thresholds				
L.ln(wage)	0.2223***	0.2986***	…	…
	(0.0575)	(0.0746)		
Overweight	0.0658	-0.0168	-0.0555***	-0.0069
	(0.0549)	(0.0342)	(0.0208)	(0.0285)
L.overwt	-0.0931**	0.0151	…	-0.0667**
	(0.0436)	(0.0297)		(0.0286)
Severely Obese	0.0155	-0.0096	0.0323	0.0361
	(0.0612)	(0.0569)	(0.0286)	(0.0345)
L.(Severely Obese)	-0.0503	-0.0234	…	0.0091
	(0.0661)	(0.0391)		(0.0363)
AR(1): *z*-statistic	-4.907	-4.891	10.07	10.11
p-value	< 0.001	< 0.001	< 0.001	< 0.001
AR(2): *z*-statistic	-0.405	-0.536	7.57	7.58
p-value	0.686	0.592	< 0.001	< 0.001
Number of Instruments	187	86	…	…
Hansen test (*p*-value)	0.947	0.818	…	…
Alternative BMI Thresholds				
L.ln(wage)	0.2197***	0.3009***	…	…
	(0.0528)	(0.0730)		
BMI ≥ 24.5	0.0177	0.0084	-0.0564***	0.0108
	(0.0852)	(0.0277)	(0.0205)	(0.0300)
L.(BMI ≥ 24.5)	-0.1039**	0.0195	…	-0.0901***
	(0.0431)	(0.0332)		(0.0298)
BMI ≥ 37	-0.1305**	0.0615	0.0419	0.0151
	(0.0564)	(0.0771)	(0.0323)	(0.0385)
L.(BMI ≥ 37)	-0.0990*	-0.0099	…	0.0525
	(0.0577)	(0.0825)		(0.0395)
AR(1): *z*-statistic	-5.028	-4.904	10.06	10.12
p-value	< 0.001	< 0.001	< 0.001	< 0.001
AR(2): *z*-statistic	-0.606	-0.604	7.58	7.59
p-value	0.545	0.546	< 0.001	< 0.001
Number of Instruments	187	86	…	…
Hansen test (*p*-value)	0.976	0.818	…	…

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 . There are 3,154 observations for 1,060 women. Specifications are as described previously, except for BMI variables being treated as exogenous in the later three columns.

Table 6. Effects of Health Shocks on Identification in the Preferred Models

	White Men			White Women		
	Preferred	Changes in Health		Preferred	Changes in Health	
	Specification	*t* only	*t* & *t*-1	Specification	*t* only	*t* & *t*-1
WHO BMI Thresholds						
L.ln(wage)	0.0720*	0.0709*	0.0707*	0.2223***	0.2255***	0.2235***
	(0.0412)	(0.0416)	(0.0414)	(0.0575)	(0.0569)	(0.0568)
Overweight	-0.1190	-0.1266	-0.1239	0.0658	0.0783	0.0954
	(0.1207)	(0.1148)	(0.1168)	(0.0549)	(0.0801)	(0.0898)
L.Overweight	0.0549	0.0569	0.0561	-0.0931**	-0.0881**	-0.0845*
	(0.0470)	(0.0477)	(0.0476)	(0.0436)	(0.0429)	(0.0459)
Severely Obese	0.0557	0.0311	0.0310	0.0155	0.0256	0.0094
	(0.0925)	(0.0991)	(0.1027)	(0.0612)	(0.0683)	(0.0700)
L.(Severely Obese)	-0.1722***	-0.1708***	-0.1697***	-0.0503	-0.0605	-0.0667
	(0.0594)	(0.0645)	(0.0643)	(0.0661)	(0.0671)	(0.0673)
Hansen *J* statistic, χ^2(df)	130.3	129.2	129.6	112.2	114.6	113.9
p-value	0.80	0.82	0.82	0.95	0.93	0.93
BMI diff-in-Hansen, χ^2(df	102.1	101.2	101.6	81.31	83.55	82.66
p-value	0.74	0.76	0.75	0.96	0.94	0.95
Alternative BMI Thresholds for White Women						
L.ln(wage)	…	…	…	0.2197***	0.2191***	0.2285***
				(0.0528)	(0.0541)	(0.0535)
BMI ≥ 24.5	…	…	…	0.0177	0.0218	0.0190
				(0.0852)	(0.1087)	(0.1021)
L.(BMI ≥ 24.5)	…	…	…	-0.1039**	-0.0993**	-0.0950**
				(0.0431)	(0.0423)	(0.0456)
BMI ≥ 37	…	…	…	-0.1305**	-0.1405**	-0.1499**
				(0.0564)	(0.0576)	(0.0619)
L.(BMI ≥ 37)	…	…	…	-0.0990*	-0.1084*	-0.1005*
				(0.0577)	(0.0554)	(0.0564)
Hansen *J* statistic, χ^2(df)	…	…	…	105.3	106.8	108.7
p-value				0.98	0.97	0.96
BMI diff-in-Hansen, χ^2(df	…	…	…	75.1	76.1	77.6
p-value				0.98	0.98	0.97

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. There are 5,897 observations for 1,473 men, and 3,154 observations for 1,060 women. General Health is reported on a 5-point scale from "Excellent" (1) to "Poor" (5), with "Excellent" being the excluded category. χ2 degrees of freedom for the Hansen *J* statistic and BMI diff-in-Hansen are 145 and 112 for men, and 136 and 103 for women.

Table A1. Additional Summary Statistics for White Men

	Mean	Std. Dev.	Min	Max
ln(Wage)	2.332	0.619	-3.252	10.081
ln(Wage) Difference	0.052	0.587	-6.202	6.142
Decrease in BMI cat.	0.031	0.174	0	1
Increase in BMI cat.	0.058	0.235	0	1
South	0.319	0.466	0	1
Urban	0.703	0.457	0	1
Part Time	0.071	0.257	0	1
Married	0.231	0.422	0	1
Any Children	0.272	0.445	0	1
HS	0.362	0.480	0	1
Some College	0.239	0.427	0	1
College	0.171	0.376	0	1
Local Unempl. Rate	6.314	2.734	0	27.8
Occupations				
Service	0.155	0.362	0	1
Mgmt, Tech., & Prof.	0.177	0.382	0	1
Sales	0.109	0.312	0	1
Clerical, Admin.	0.088	0.284	0	1
Misc. Blue Collar	0.470	0.499	0	1

Note: As in Table 1, there are 9,037 observations used for most of these variables. Occuption summarized where not missing. "Decrease (or Increase) in BMI cat." refers to the fraction who move to a lower (or higher) BMI category as defined by the WHO.

Table A2. Additional Summary Statistics for White Women

	Mean	Std. Dev.	Min	Max
ln(Wage)	2.196	0.563	-3.069	6.652
ln(Wage) Difference	0.044	0.488	-6.060	4.499
Decrease in BMI cat.	0.026	0.159	0	1
Increase in BMI cat.	0.051	0.220	0	1
South	0.331	0.470	0	1
Urban	0.757	0.429	0	1
Part Time	0.115	0.319	0	1
Married	0.253	0.435	0	1
Any Children	0.204	0.403	0	1
HS	0.280	0.449	0	1
Some College	0.282	0.450	0	1
College	0.319	0.466	0	1
Local Unempl. Rate	6.237	2.674	0	19.2
Occupations				
Service	0.236	0.425	0	1
Mgmt, Tech., & Prof.	0.314	0.464	0	1
Sales	0.133	0.340	0	1
Clerical, Admin.	0.248	0.432	0	1
Misc. Blue Collar	0.069	0.254	0	1

Note: As in Table 1, there are 5,408 observations used for most of these variables. Occuption summarized where not missing. "Decrease (or Increase) in BMI cat." refers to the fraction who move to a lower (or higher) BMI category as defined by the WHO.