



Munich Personal RePEc Archive

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

Pinkston, Joshua

University of Louisville

8 May 2015

Online at <https://mpra.ub.uni-muenchen.de/75543/>  
MPRA Paper No. 75543, posted 12 Dec 2016 05:49 UTC

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

Joshua C. Pinkston  
University of Louisville

October 12, 2016

## 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](mailto: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.<sup>1</sup>

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,

---

<sup>1</sup>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 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. The only 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. The current paper provides a detailed look inside of those early years using a more recent cohort of workers.<sup>2</sup>

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.<sup>3</sup> Odgen et al. (2010) report that obesity doubled among adults in the U.S. between 1980 and 2000. If prejudice or stereotypes evolve

---

<sup>2</sup>Another difference between Chen (2012) and the current paper is the treatment of endogeneity. Chen (2012) notes that adding future body mass to her wage regressions suggests a problem with endogeneity, but she does not otherwise address the problem.

<sup>3</sup>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).

as the population becomes heavier, results from previous work based on the NLSY79 may not generalize to more recent cohorts.<sup>4</sup>

Finally, the estimation sample is unique in its focus on young workers in the first several years after entry into the labor market. As a result, the estimation should be better able to capture potential discrimination as it unfolds than it would with a sample of older workers. Wage growth is higher earlier in careers, and changes in jobs and occupations are more important.<sup>5</sup> Removing fixed effects in a sample of older workers is likely to remove the accumulated effects of past discrimination suggested by Chen (2012). Furthermore, signals inferred from body mass should have larger effects for younger than for older workers because the market knows less about younger workers.<sup>6</sup>

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

---

<sup>4</sup>Additionally, Altonji et al. (2012) find that the mix of skills and family backgrounds changed between NLSY cohorts, which would affect the ability to generalize across decades.

<sup>5</sup>See Murphy and Welch (1992), Topel and Ward (1992), and Neal (1999) among others.

<sup>6</sup>Hamermesh (2011) provides a similar argument to explain why effects of beauty on wages might decline with age. See Altonji and Pierret (2001) for more on statistical discrimination in the presence of employer learning. Lange (2007) discusses how quickly the market learns about young workers.

Employer learning about healthcare costs may be more complicated than learning about productive characteristics if the relationship between BMI and expected healthcare costs changes with age.

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

$$w_{it} = X_{it}\beta + BMI_{it}\phi + \nu_{it}, \quad (1)$$

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,  $\nu_{it}$ , 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.<sup>7</sup> Other estimates used instrumental variables to address the bias in equation (1); however, the instruments

---

<sup>7</sup>E.g., Cawley (2004), Baum and Ford (2004), and Han et al. (2009).

used so far in the literature are correlated with individual fixed effects, and likely correlated with relevant time-invariant unobservables.<sup>8</sup> 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.<sup>9</sup> This assumption is inconsistent with the broader discrimination literature in which dynamic models are now common. For example, Lang et al. (2005) show that hiring decisions in a wage posting model magnify effects of prejudice or expected productivity differences, resulting in wage differences even if employers are not willing to pay more to hire their preferred group.<sup>10</sup> Bjerck (2008) shows that any one of a few types of statistical discrimination can result in wage differentials over time due to effects on hiring and promotions.<sup>11</sup> Adapting either of these models to discrimination based on weight would imply that a worker's history of body mass could affect her current wages through effects on past job search and promotions. Furthermore, any

---

<sup>8</sup>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 the NLSY97, even after 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 even a single gene can affect multiple relevant outcomes.

<sup>9</sup>As noted above, Chen (2012) is a recent, important exception.

<sup>10</sup>Differences in healthcare costs should have similar effects to productivity differences in Lang et al. (2005), which would fit the results of Bhattacharya and Bundorf (2009).

<sup>11</sup>Additionally, Oettinger (1996) shows that differences in uncertainty about workers' productivity can affect workers' returns to mobility.

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.

Regardless of how or why wage penalties arise, the idea that the full effects of changes in body mass would be captured by immediate changes in the wage is inconsistent with the presence of labor-market frictions. 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.<sup>12</sup>

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 past wages.

---

<sup>12</sup>See Kahn (2010) and Oreopoulos et al. (2012). Plant closings and layoffs have also been found to have persistent negative effects on wages. See, e.g., Jacobson et al. (1993).



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

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

Both  $BMI_{it}$  and  $BMI_{it-1}$  are potentially correlated with the individual fixed effect,  $\alpha_i$ , as is  $w_{it-1}$ .<sup>13</sup> As before,  $BMI_{it}$  may be correlated with the time-varying error,  $\varepsilon_{it}$ , 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).<sup>14</sup> The first step in this approach is to use differences to eliminate the fixed effect:

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

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

Fortunately, further lagged levels of the wage are valid instruments for  $\Delta w_{it-1}$  if there is no serial correlation in  $\varepsilon$ . Under this assumption,  $w_{it-2}$  is not correlated with  $\varepsilon_{it}$  or  $\varepsilon_{it-1}$ , but is correlated with  $\Delta w_{it-1}$ .<sup>15</sup> The GMM estimator of HENR and AB also uses further lags, where available, as

---

<sup>13</sup>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.

<sup>14</sup>See Arellano and Honoré (2001), Bond (2002), or Arellano (2003) for reviews.

<sup>15</sup>The correlation of  $w_{it-2}$  and  $\Delta w_{it-1}$  is weak if  $\gamma$  is close to 1; however, the results presented below suggest this is not a problem.

instruments to improve efficiency.

It is important to note that assuming no serial correlation in the time-varying errors,  $\varepsilon_{it}$ , 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  $\varepsilon$  is violated only if there is serial correlation in residual heterogeneity that is uncorrelated with lagged wages, BMI (lagged and current), and any other regressors included in  $X_{it}$ .<sup>16</sup>

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  $\Delta BMI_{it}$  and  $\Delta BMI_{it-1}$  as long as  $BMI_{it}$  is uncorrelated with  $\varepsilon_{it+1}$  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  $\varepsilon_{it}$ , this assumption requires that those shocks only affect  $w_{it+1}$  through their effects on  $BMI_{it}$  and  $w_{it}$ .<sup>17</sup>

Finally, the identification of equation (3) requires changes in BMI to be

---

<sup>16</sup>In contrast, serial correlation in the residuals of static wage regressions, which dominate the previous literature, 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.

<sup>17</sup>Shocks to  $BMI_{it}$  that are correlated with both  $\varepsilon_{it}$  and  $\varepsilon_{it+1}$  would imply serial correlation in  $\varepsilon$ , which is testable. It is possible that some shocks to  $BMI_{it}$  are uncorrelated with  $\varepsilon_{it}$  but correlated with  $\varepsilon_{it+1}$ , but this is also testable in many cases. Section 3.1 discusses tests of identifying assumptions in light of such possibilities. Section 4 presents results of these tests, as well as additional falsification tests.

predicted by its lagged levels.  $BMI_{it-2}$  should be correlated with  $\Delta BMI_{it}$  if  $BMI_{it}$  is endogenous.<sup>18</sup> This assumption finds empirical support in the dynamic models of body mass estimated by Goldman et al. (2010) and Ng et al. (2010). Studies in the epidemiology literature also find that large changes in weight are more common among those who were initially heavier.<sup>19</sup> 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  $\Delta BMI_{it}$ , with  $F$  statistics above traditional cutoffs for weak instruments.<sup>20</sup>

## 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.

---

<sup>18</sup>If  $BMI_{it}$  were merely predetermined, meaning it was correlated with  $\varepsilon_{it-1}$  but not  $\varepsilon_{it}$ ,  $BMI_{it-1}$  would also be a valid instrument for  $\Delta BMI_{it}$ .

<sup>19</sup>E.g., see Lewis et al. (2000) and Williamson et al. (1990).

<sup>20</sup>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.

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 they were not asked about weight in '79, '80, '83, '84 or '87. As a result, NLSY79 respondents were between 25 and 33 years old in the first year (1990) that BMI 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 valid wage observations reduces the number of observations more for minorities and women than for white men.

Additionally, only jobs following full-time labor market entry are included in the estimation sample. Entry is defined as the first two consecutive years in which the individual works full time and is employed 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 adult careers.<sup>21</sup> 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

---

<sup>21</sup>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.

as observations for women who were pregnant at any point since the last interview. Limiting attention to observations that can be use as time  $t$ ,  $t - 1$ , or  $t - 2$  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.<sup>22</sup> 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.<sup>23</sup> All regressions that include current or lagged BMI variables also include corresponding dummy variables for interviews being conducted by phone.

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

---

<sup>22</sup>BMI is defined as  $(\text{weight in kg})/(\text{height in m})^2$

<sup>23</sup>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.

NHANES samples as validation data for NLSY cohorts are not credible given the mix of interview methods in the NLSY.<sup>24</sup> Secondly, wage penalties are more likely driven by whether the worker’s appearance conforms to desired standards than by an actual BMI number, which is unlikely to be observed by employers. Even measured BMI, therefore, should not be seen as the “true” variable of interest in studies of labor market outcomes.<sup>25</sup> Finally, 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.

## 2.2 The Estimation Sample

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. 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.<sup>26</sup> In 2009,

---

<sup>24</sup>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.

<sup>25</sup>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 his use of digitally manipulated photographs to estimate the effects of appearing more or less fat.

<sup>26</sup>Respondents were as young as 16 in period  $t - 2$ , or 18 in  $t$ . This introduces the possibility that changes in BMI mostly reflect changes in height for part of my sample

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 ( $\text{BMI} \leq 18.5$ ). Over 57% of white men and nearly 42% of white women are overweight ( $\text{BMI} \geq 25$ ). 22% of white men and 20% of white women are obese ( $\text{BMI} \geq 30$ ). Almost 8% of men and over 10% of women report a BMI that qualifies as severely obese ( $\text{BMI} \geq 35$ ).

As seen in the appendix tables, movement between official BMI categories is not uncommon.<sup>27</sup> 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.<sup>28</sup> Therefore, it does not appear as though identification is comes from small fluctuations in weight that employers would not notice.

---

and changes in weight for the rest; however, the results presented below are robust to excluding respondents under the age of 20.

<sup>27</sup>Roughly 9% of men and 8% of women move from one category to another between two consecutive years. If we consider three consecutive years, roughly 15% of men and 13% of women in our estimation sample have moved between categories at least once (not shown).

<sup>28</sup>The median percent change in BMI associated with a change in BMI category is 8.5 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.

### 3 Estimation

Recall that the differenced equation we're interested in takes the form

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

As mentioned in the first section, the GMM approach developed by HENR and AB uses second and further lagged levels as instruments for  $\Delta w_{it-1}$  and other endogenous variables. The GMM instruments enter as separate vectors for each year.<sup>29</sup> If a lag is missing, it enters as a zero. This allows the use of further lags without limiting observations to cases that have those lags.

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

#### 3.1 Testing Assumptions

In Section 1, I assumed that the time-varying errors,  $\varepsilon_{it}$ , 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

---

<sup>29</sup>At  $t = 3$ ,  $w_{i1}$  is an instrument for  $\Delta w_{i2}$ . At  $t = 4$ ,  $w_{i2}$  and  $w_{i1}$  are instruments for  $\Delta w_{i3}$ , and so on.

<sup>30</sup>See Roodman (2006) for documentation. Stata contains other dynamic panel-data programs, but they do not provide the same array of tests as XTABOND2.



$\varepsilon$ , equation (3) will be at least  $AR(2)$ . I present tests for serial correlation with all of the results that follow.<sup>31</sup>

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  $\varepsilon_{it}$  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 further lags are valid instruments.<sup>32</sup>

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  $\varepsilon_{it+1}$  but not  $\varepsilon_{it+2}$ , the second lags would not be valid instruments but the third lags would be. Overidentification based on lagged BMI variables would be theoretically untestable only if  $BMI_{it}$  was correlated with  $\varepsilon_{it+1}$ ,  $\varepsilon_{it+2}$ ,  $\varepsilon_{it+3}$ , etc.; however, the absence of serial correlation in  $\varepsilon$  would make this unlikely.

---

<sup>31</sup>I performed tests  $AR(3)$  and higher, but I only present results for  $AR(2)$ . Results for  $AR(3)$  and higher are available on request.

<sup>32</sup>The Arellano-Bond tests would have to miss a lot of serial correlation 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  $\Delta w_{it-1}$ , 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.

For example,  $BMI_{it}$  could be correlated with  $\varepsilon_{it+1}$  and later residuals due to health shocks that affect body mass more immediately than they affect wages. But such health shocks would imply serial correlation in  $\varepsilon$ , unless the unobserved factors that were common to  $BMI_{it}$  and  $\varepsilon_{it+1}$  were somehow independent of the factors common to  $BMI_{it}$  and  $\varepsilon_{it+2}$ . Although this strikes me as improbable, it is always possible that tests of overidentification or serial correlation are too weak in practice to reject a hypothesis that theoretically could be rejected.<sup>33</sup> Therefore, I consider the possible effects of health shocks on identification when I discuss the robustness of results in Section 4.2.

### 3.2 Other Potential Problems with Instruments

The use of all lagged levels of wage and BMI variables quickly 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 presented in the next section 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.

---

<sup>33</sup>My goal here and elsewhere is to convince readers that my identifying assumptions are reasonable and weaker than they may seem at first glance. I acknowledge that the identification of causal effects in this paper, as in any other, relies on assumptions.

### 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.<sup>34</sup> 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 that

---

<sup>34</sup>In preliminary estimation, linear and polynomial specifications of  $BMI_{it}$  only produced statistically significant results in static OLS regressions. Ignoring statistical significance, the results based on cubic polynomials are roughly consistent with the preferred dummy-variable specifications.

range from 23 to 38.<sup>35</sup> This allows me to compare the robustness of coefficient estimates for a BMI category to other changes in the specification of  $BMI_{it}$ . I also compare similar specifications using the tests proposed by Bond et al. (2001) and Andrews and Lu (2001), which are both based on comparisons of the Hansen  $J$  statistic.<sup>36</sup> Finally, I consider the robustness of my preferred specifications to the treatment of outliers in the distribution of wages.<sup>37</sup> In the end, I find that specifications based on traditional cutoffs perform surprisingly well.

All of the dynamic models presented below include only one lag of wage and one lag of  $BMI$ . None of the tests for serial correlation suggest that more lags of the wage 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 qualitatively change the results; however, adding further lags may simply ask too much of the data. The only noticeable effect of adding a second lag of either BMI or the wage is a reduction of observations by 20% for men and 25% for women.<sup>38</sup>

All regressions control for the local unemployment rate and incidence of obesity in the state, as well as dummy variables for region, urban residence,

---

<sup>35</sup>I estimated models with three dummy variables, but they provided no obvious advantage over those with two.

<sup>36</sup>I used leave-one-out crossvalidation, dropping all observations for each respondent one at a time, to ensure that these test results were not peculiar to the estimation sample.

<sup>37</sup>As noted in the appendix, I only drop wages if  $\Delta \ln(w_{it})$  is greater than 6.5 in absolute value. This leaves some observations with unusually high or low wages.

<sup>38</sup>An advantage of focusing on young workers is that they have less history in the market to control for. In a sample of older workers, one lag of wages or BMI would be expected to capture less of a worker's relevant past.

and being interviewed by telephone.<sup>39</sup> When lagged values of BMI variables are included as regressors, 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 the number of years since labor market entry. No estimation in this paper is weighted.

I also control for actual experience in the labor market and its square in the dynamic panel estimates. 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.<sup>40</sup> Since actual experience may be endogenous, potential experience (age—schooling—six), its square and its interactions are used as traditional instruments in the dynamic models. In OLS regressions, I simply replace actual experience with potential.

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. After presenting the main results, I examine the robustness of those results to the addition of these and other variables. I discuss the possibility that some of these variables may be intermediate outcomes through which body mass affects wages, as well as possible implications for

---

<sup>39</sup>Local unemployment and state identifiers are provided by the NLSY97 Geocode files. Percent obese in the state is estimated by the CDC.

<sup>40</sup>Lagged wages might reflect the accumulation of experience or commitment to the labor market. Overidentification tests for lagged wage instruments improve with the use of actual experience and its interactions.

my identifying assumptions.

## 4 Results

I find that white men face a penalty for past severe obesity, while white women face a penalty for being overweight at all in the past year. Considering alternative BMI thresholds, women are penalized for a past BMI over 24.5, with additional penalties for past BMI of 37 or more. The only evidence I find of an effect of current BMI on wages is a penalty faced by women with a reported BMI of at least 37. Additionally, the results support the use of autoregressive wage equations and suggest that BMI is endogenous even after removing individual fixed effects.

The next subsection discusses the main results, and then compares the preferred dynamic models to simpler regressions. Section 4.2 considers the robustness of the main results to several potential sources of bias.

### 4.1 The Preferred Models

As discussed in section 3.3,  $BMI_{it}$  is modeled using dummy variables for exceeding various BMI thresholds. For white men and then white women, I first present specifications based on the familiar categories of overweight, obese and severely obese before I discuss 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.<sup>41</sup> 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 were serial correlation in the residuals.<sup>42</sup> 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.<sup>43</sup>

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

---

<sup>41</sup>All specifications in Tables 2 and 3 use lagged indicators of overweight, obesity and severe obesity as instruments. This ensures that the instruments capture the same amount of variation in each specification. 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.

<sup>42</sup>I also found no evidence of the residuals being  $AR(3)$  or higher.

<sup>43</sup>Results for the model that specified  $BMI_{it}$  using overweight and obese are excluded. They are consistent with those in columns 2 and 3.

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 robust.<sup>44</sup> There is no robust evidence of an effect at lower levels of current or lagged body mass for white men.<sup>45</sup> Furthermore, the specification test proposed by Andrews and Lu (2001) supports models that use severe obesity over similar models that use nearby cutoffs.<sup>46</sup>

I select a preferred specification from Table 2 using the test of parameter restrictions proposed by Bond et al. (2001).<sup>47</sup> The only restricted specifica-

---

<sup>44</sup>The coefficient on lagged severe obesity is qualitatively similar and statistically significant in specifications with lower cutoffs ranging from 24 to 33. Coefficients on lagged  $BMI \geq 34.5$  are similar, but smaller and statistically significant in fewer models.

<sup>45</sup>Coefficients on  $BMI \geq 29.5$  are statistically significant in a few specifications, but are not robust to the treatment of outliers. They are driven by a single observation in which the reported wage fell from \$7.11 to \$0.04.

<sup>46</sup>Andrews and Lu (2001) develop a GMM analogue of the Bayesian Information Criterion. When comparing models that have the same number of variables and instruments, their approach amounts to selecting the model that minimizes the Hansen  $J$  statistic. For example, the  $J$  statistics of models that replace severe obesity in column 5 of Table 2 with indicators for  $BMI \geq 34.5$  or  $35.5$  rise to 135.5 and 139.3, respectively.

<sup>47</sup>Bond et al. (2001) show that differences between the  $J$  statistics of restricted and unrestricted specifications are distributed  $\chi^2$  with degrees of freedom equal to the number of restrictions.



tion that cannot be rejected in favor of the full specification (column 1) is the specification in column 5, which models  $BMI_{it}$  using dummy variables for being overweight and severely obese.<sup>48</sup> The Andrews-Lu test also selects this specification over the others that use two dummy variables to model  $BMI_{it}$ . 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 as the preferred specification for white men.

Using alternative BMI thresholds may be more important for white women than for white men. The penalty for a lagged  $BMI \geq 24.5$  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 estimated coefficients on lagged  $BMI \geq 24.5$  (columns 3-5) are similar, but consistently larger than analogous coefficients

---

<sup>48</sup>Columns 2 and 3 are rejected at any conventional level. Column 6 is rejected at a 5% level, and column 4 at a 10% level.

on lagged overweight status (columns 1 and 2).<sup>49</sup>

When I consider alternative thresholds in the heavier part of the BMI distribution, I find some evidence of an additional penalty for a BMI  $\geq 37$ . This corresponds to a penalty for being in the heaviest eight percent of the distribution, as does the penalty faced by severely obese white men.<sup>50</sup> The results, shown in columns 2 and 5 of Table 4, suggest a penalty for a current BMI of 37 or more of 13 to 14 percent. Furthermore, the coefficient on a lagged BMI  $\geq 37$  is -0.099 (0.038) in column 6.<sup>51</sup>

The Hansen  $J$  statistics in Table 4 support the use of reported overweight status instead of the slightly lower threshold of 24.5; however, the direction of this difference is not robust to the treatment of outliers or to small changes in the set of instruments used.<sup>52</sup> In either case, the results suggest that women who reported a BMI at or slightly below the threshold for overweight in the previous year face a penalty of 9-12 percent. Considering the distribution of BMI in my sample, these results suggest that over 40 percent of young white women face some penalty for their weight.

---

<sup>49</sup>Coefficients on lagged overweight status are statistically significant in roughly half as many models as coefficients on lagged BMI  $\geq 24.5$ .

<sup>50</sup>7.9% of women in the sample report a BMI  $\geq 37$ , while 7.8% of men report being severely obese.

<sup>51</sup>The number of instruments changes when a higher threshold of 37 is used instead of the lower cutoff for severe obesity of 35, making a comparison of Hansen  $J$  statistics between specifications difficult.

Coefficients for BMI cutoffs of 36.5 and 37.5 are similar to those for BMI  $\geq 37$ , but are smaller in magnitude and statistically significant in fewer alternative models.

<sup>52</sup>Specification that use BMI  $\geq 24.5$  have smaller  $J$  statistics if dummy variables for lags of obesity are not included in the set of instruments, or if lagged overweight status is used as an instrument in both specifications. In contrast, such changes in the instrument vector have no effect on the relative performance of specifications for men.

Regardless of specification, I find that lagged wages have larger effects on current wages for white women than for white men. The coefficients on lagged log wages range from 0.211 (0.057) to 0.231 (0.057) for women, compared to 0.077 (0.042) for men.<sup>53</sup> 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 cast doubt on the validity of instruments in the autoregressive models of Tables 2 through 4.

On a more meaningful level, autoregressive wage equations imply that workers' history of body mass affects wages beyond the single lags included in the model. The wage in  $t - 1$  is a function of  $BMI_{it-2}$ . It's also a function of  $w_{it-2}$ , which is affected by  $BMI_{it-3}$ , and so on. This is an important advantage of dynamic models: In contrast to previous work, the models in this paper do not assume that the penalty for being heavy is the same as the penalty for becoming heavy. Wages are allowed to respond slowly to changes in BMI, implying 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

---

<sup>53</sup>Note that the weak identification problem discussed by Blundell and Bond (1998) is only relevant when the coefficients on lagged wages approach one.

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%.<sup>54</sup> 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 initially faces a penalty of roughly 13%, according to the model in column (5). 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  $t - 1$ . Women with a reported BMI at (or slightly below) the threshold for being overweight face a penalty of 9 to 12 percent after one year that approaches 12 to 16 percent over time.

#### 4.1.3 Comparisons to Simpler Models

Tables 5A and 5B compare the preferred dynamic specifications to OLS regressions and an autoregressive model that assumes  $\Delta BMI_{it}$  is exogenous. The OLS results for men in Table 5A suggest that overweight men are paid

---

<sup>54</sup>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  $\Delta w_{it-1}$ . If the woman was also overweight in  $t - 1$ , her combined penalty in  $t$  is -0.148 (0.056).

more than lighter peers while severely obese men are paid less.<sup>55</sup> The OLS results in Table 5B suggest that being overweight (upper panel) or having a  $BMI \geq 24.5$  (lower panel) is associated with lower wages for white women, which is consistent with previous research; however, this negative association is driven by lagged BMI, not current.

The autoregressions presented in the second columns of Tables 5A and 5B remove any bias that may arise from correlation with individual fixed effects or the omission of lagged wages, but treat  $\Delta BMI_{it}$  and  $\Delta BMI_{it-1}$  as exogenous. Tests of overidentifying restrictions (not shown) reject this assumption, suggesting that changes in BMI are correlated with time-varying unobservables.<sup>56</sup> 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 static OLS regressions shown in Tables 5A and 5B are serially correlated, while the errors 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 equations. In contrast, the dynamic wage equa-

---

<sup>55</sup>When lagged BMI variables are added, the OLS coefficients on current and lagged dummies for BMI categories are jointly significant for white men.

<sup>56</sup>I reject the exogeneity of the  $BMI_{it}$  for white men when I use only the second and third lags of  $w$  as instruments. (Using more instruments weakens overidentification tests.) I reject the exogeneity of  $BMI_{it}$  for women in models that include further lagged levels of  $BMI$  as instruments.

tions explicitly model the correlation between past and current wages that creates serial correlation in the residuals of 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

### 4.2.1 Robustness to Additional Control Variables

There are a number of variables that might be correlated with both wages and BMI that have been excluded so far. Some of these variables are potential intermediate outcomes through which body mass could influence wages. Examining the robustness of results to the inclusion of such variables may shed light on potential mechanisms connecting BMI to wages. Furthermore, as noted by Angrist and Krueger (1999) among others, sensitivity of the results to the inclusion (or exclusion) of different observable characteristics could suggest sensitivity to the exclusion of unobservable characteristics.

The results presented above are surprisingly robust to the addition of potentially endogenous confounders, as well as other changes in the set of control variables.<sup>57</sup> The first such variables I consider are marital status, number of children, and indicators for age of the youngest child being less than or equal to two or five. There are a number of reasons these variables may be correlated with wages and BMI, especially for women; however, their addition to my preferred models has no noticeable affect on the estimated

---

<sup>57</sup>All of the results discussed in this subsection are available on request; however, I do not present most of them for the sake of brevity.

coefficients.<sup>58</sup>

One might expect a high BMI to be related to hours worked, the duration of employment relationships or investment in on-the-job training; however, I find no evidence that these are important confounders or intermediate outcomes. I examined hours worked by adding current and lagged dummy variables for working part time or working 60 hours or more per week. I also added current and lagged levels of tenure, as well as indicators for the receipt of on-the-job training.<sup>59</sup> In each case, the estimated coefficients of interest were similar to those presented above.

I also find no evidence of BMI affecting wages through selection into jobs that offer health insurance. Working for an employer that offers health insurance is associated with higher wages. A wage differential could arise in part from heavier workers being pushed out of jobs that offer insurance; however, adding indicators for employer-provided insurance has little effect on the results.<sup>60</sup>

## Occupation

As shown in Table 6, indicators for occupation are the only variables I've found that do affect the results of my preferred models. The estimated

---

<sup>58</sup>For discussions of BMI and marriage, see Averett et al. (2008) or Chiappori et al. (2012). The presence and age of children would more likely be an endogenous confounder than an intermediate outcome.

<sup>59</sup>I considered specifications of training as both a dummy variable for training in the past year, and a count of years in which training was received.

<sup>60</sup>This result does not directly contradict Bhattacharya and Bundorf (2009), who suggest that wage penalties vary with whether health insurance is offered or not. Unfortunately, extending my estimation to investigate their story is not practical.

effects of BMI on the wages of women decrease in magnitude when controls for occupation are added. Only the coefficient for a lagged  $\text{BMI} \geq 24.5$  is still statistically significant at even a 10% level, and it falls from -0.103 (0.043) to -0.078 (0.046). In contrast, the effect of past severe obesity on the wages of white men rises, if anything.

The regressions presented in Table 6 use second lagged levels of occupation as instruments because the exogeneity of current occupation can be rejected for both genders. Although the second lags cannot be jointly rejected at conventional levels for either gender, the  $p$ -value for women are not reassuring. Furthermore, the second lag of managerial, professional and technical occupations can be rejected at conventional levels for women. Especially when combined with the changes in coefficients on BMI variables, these results are consistent with those in Lakdawalla and Philipson (2007), which suggest that body mass drives selection into occupations for women.<sup>61</sup>

### **Health Shocks, Ability & Identification**

The identifying assumptions in this paper already received an unusual amount of empirical scrutiny in Section 4.1; however, further examination is justified for two reasons. 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.<sup>62</sup> Secondly, as discussed in Section 3.1, the validity of

---

<sup>61</sup>Also see Han et al. (2011), and the recent structural work of Harris (2015). Like Lakdawalla and Philipson (2007), these papers use data from the NLSY79.

<sup>62</sup>Fortunately, tests of serial correlation and overidentifying restrictions do reject hypotheses that we expect to be rejected, but type II errors are still possible.



using lagged levels of  $BMI$  as instruments would be theoretically untestable if  $BMI_{it}$  were somehow correlated with all future wage residuals ( $\varepsilon_{it+1}$ ,  $\varepsilon_{it+2}$ , etc.) without those residuals being correlated with each other.

The most obvious reason why  $BMI_{it}$  might 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  $BMI_{it}$ . Although health shocks that cause  $BMI_{it}$  to be correlated with both  $\varepsilon_{it+1}$  and  $\varepsilon_{it+2}$  would likely cause serial correlation in  $\varepsilon$ , which I don't find, I examine the possibility that my main results are biased by such shocks in Table 7.

Table 7 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 correlated with body mass as one would expect.<sup>63</sup> 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 7 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

---

<sup>63</sup>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.

recent changes or also include lagged changes in health, the coefficients on  $BMI$  are similar.<sup>64</sup> Furthermore, none of the overidentification tests at the bottom of each panel change in a way that suggests health shocks affect the identification of the BMI variables, or the overall identification of the model.

Another reason  $BMI_{it}$  may be correlated with future wage residuals is that both may be affected by individual ability. Either human capital accumulation or employer learning could result in time-invariant individual ability having effects on wages that vary over time. If ability is also correlated with maintaining a desirable body mass, the results in this paper could be biased, despite the use of first-differences.<sup>65</sup>

I examine this possibility by adding the interaction of AFQT with time in the labor market to the preferred models. If time-varying effects of time-invariant ability bias my results, the results should not be robust to allowing an ability measure to affect wage growth. Adding AFQT interactions, however, have no discernible effect on the coefficient estimates or the test statistics presented in previous tables.<sup>66</sup>

---

<sup>64</sup>I also estimated models with up to five lags of the health indicators, and models that use the lagged residuals of  $BMI$  regressed on health indicators as instruments. In each case, the results were similar to those in Table 7.

The results for  $BMI$  are insensitive to assumptions about the exogeneity of general health, but the results for health indicators (not shown) are. “Very good” health is associated with higher wages than “excellent” health, but only when reported health is treated as exogenous.

<sup>65</sup>In Section 1, I questioned the validity of sibling BMI as an instrumental variable on the grounds that it predicts AFQT, even if we control for the respondent’s BMI. The differencing used in the current paper rules out bias from time-invariant effects of ability, but not time-varying effects of ability.

<sup>66</sup>Results are available on request. The coefficient on AFQT interacted with years in the market is not statistically significant in any specification.

## 4.2.2 Sample Selection

There are two reasons to worry about selection into the estimation sample biasing results in this paper: The dynamic models require three consecutive years of labor market participation, and I've limited attention to workers who have worked full time for at least 75% of two consecutive years. On the other hand, the differenced estimator is only affected by selection bias if selection into the sample is based on time-varying unobservables. Bias from selection on unobservables that are constant over time is removed 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  $t - 2$ .<sup>67</sup> I then add the resulting inverse Mills ratios and their interactions with time dummies to the wage regressions. The null hypothesis of no selection is rejected if coefficients on the Mills ratios and their interactions are jointly significant.

These tests do not indicate that the main results of the paper are biased by selection on time-varying unobservables. None of the coefficients on the inverse Mills ratios or their interactions are statistically significant. The  $p$ -value on the test of joint statistical significance is 0.26 for white men and 0.29 for white women.

---

<sup>67</sup>AFQT is used as an instrument in the probit estimates. AFQT strongly predicts participation, but it is differenced out of the wage equations.

### 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  $\Delta BMI_{it}$ , one might still worry that there is not enough variation in the lagged levels to identify coefficients on both  $\Delta BMI_{it}$  and  $\Delta BMI_{it-1}$ .<sup>68</sup> In that case, the estimates in my preferred specifications, especially for  $\Delta BMI_{it}$ , would be biased toward zero.<sup>69</sup>

One way to address a potential lack of variation would be adding lags 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.<sup>70</sup>

I also compared the preferred specifications to restricted specifications that include either  $\Delta BMI_{it}$  or  $\Delta BMI_{it-1}$ , but not both.<sup>71</sup> If there is an effect of  $\Delta BMI_{it}$  that we can't observe in the main results due to insufficient variation in the instruments, we would expect a change in the coefficients on  $\Delta BMI_{it}$  when  $\Delta BMI_{it-1}$  is excluded.<sup>72</sup>

---

<sup>68</sup>The GMM estimator of HENR exploits a larger set of moment conditions than traditional IV estimation. Regressions of  $\Delta BMI_{it}$  on lagged levels are not analogous to the first stage of traditional 2SLS.

<sup>69</sup>More precisely, the coefficients would be biased toward those in the second columns of Tables 5A and 5B, which are smaller and statistically insignificant.

<sup>70</sup>I used only the second and third lags of the continuous variables to avoid a large increase in the number of instruments.

<sup>71</sup>Kropfhäuber and Sunder (2015), citing an early version of the current paper, estimate similar wage equations that only include BMI variables from one year. They use German data that has wage information every other year and BMI in off years. With  $t$  indexing a single year, 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.

<sup>72</sup>The coefficients could also change because  $\Delta BMI_{it}$  is correlated with  $\Delta BMI_{it-1}$ .

The results (not shown) do not suggest that the preferred specifications are affected by weak instruments. Dropping  $\Delta BMI_{it-1}$  from wage equations does not reveal effects of current BMI that were not previously identified. In fact, the estimated penalty white women face for a current  $BMI \geq 37$  is only statistically significant when lagged BMI is also included in the regression. Furthermore, comparing Hansen  $J$  statistics between specifications, as suggested by Bond et al. (2001), rejects the restricted specifications for both men and women in favor of models that include both  $\Delta BMI_{it}$  and  $\Delta BMI_{it-1}$ .

## 5 Discussion

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, in contrast to previous work, my results make it clear that the penalty for becoming heavy is not the same as the penalty for being heavy throughout the early years of a career. Penalties for a high body mass can persist and even accumulate over time, especially for women.

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 jobs with employer-provided health insurance. The main

results are robust to the addition of controls for general health or having a job that offers insurance coverage. Furthermore, the differences in penalties by gender seem more consistent with discrimination based on appearance than with impaired health or expected healthcare costs.

The results are consistent with dynamic models of discrimination that incorporate labor market frictions. For example, models in which discrimination affects labor market search, like those in Bowlus and Eckstein (2002) and Lang et al. (2005), are consistent with penalties based on recent lags of body mass. The idea that body mass can affect labor market search is also supported by the field experiment of Rooth (2009), who finds that obese job applicants receive fewer interview requests.

Finally, persistent effects of high body mass on wages are consistent with effects on occupational selection. Some of the evidence in this paper points to occupation as an intermediate outcome through which body mass affects wages, especially in the case of white women. The results suggest that the occupational selection observed in the NLSY79 by Lakdawalla and Philipson (2007) and Harris (2015) still affects 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 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.
- Bhattacharya, J. and M. K. Bundorf (2009). The incidence of the healthcare costs of obesity. *Journal of Health Economics* 28(3), 649–658.
- 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.
- 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.
- Chiappori, P.-A., S. Oreffice, and C. Quintana-Domeque (2012). Fatter attraction: anthropometric and socioeconomic matching on the marriage market. *Journal of Political Economy* 120(4), 659–695.
- 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.
- Kropfhäuser, 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., 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.
- Oettinger, G. S. (1996). Statistical discrimination and the early career evolution of the black-white wage gap. *Journal of Labor Economics* 14(1), 52–78.
- 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.
- 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 drops 3,752 respondents, leaving 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.<sup>73</sup>

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 also 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.<sup>74</sup>

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

---

<sup>73</sup>This includes 63 observations in which a respondent reported a military occupation despite not being otherwise identified as in the military.

<sup>74</sup>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.

The appendix tables A1 and A2 present summary statistics for men and women that are not presented in Table 1. The mean wage for white men is pulled up by outliers; however, estimates do not appear sensitive to dropping these high wage observations. Median hourly wages (not shown) are \$9.86 for men and \$8.98 for women.

As expected, the sample is largely 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 who are committed enough to the labor market to be 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
<i>White Men</i>				
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
-----				
<i>White Women</i>				
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* (0.0414)	0.0605 (0.0410)	0.0648 (0.0416)	0.0769* (0.0420)	0.0720* (0.0412)	0.0768* (0.0421)
Overweight	-0.1187 (0.1198)	-0.0977 (0.1224)	...	...	-0.1190 (0.1207)	...
L.Overweight	0.0540 (0.0491)	0.0477 (0.0456)	...	...	0.0549 (0.0470)	...
Obese	0.0112 (0.0979)	...	0.0054 (0.1021)	...	...	-0.0051 (0.1012)
L.Obese	0.0030 (0.0567)	...	0.0175 (0.0576)	...	...	0.0061 (0.0600)
Severely Obese	0.0465 (0.0927)	...	...	0.0745 (0.0938)	0.0557 (0.0925)	0.0360 (0.0951)
L.Severely Obese	-0.1699*** (0.0632)	...	...	-0.1649*** (0.0598)	-0.1722*** (0.0594)	-0.1668** (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
<u>Diff-in-Hansen Tests for Exogeneity of Subsets of GMM Instruments (p-values)</u>						
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.In(wage)	0.2110*** (0.0567)	0.2308*** (0.0568)	0.2158*** (0.0542)	0.2209*** (0.0557)	0.2237*** (0.0577)	0.2223*** (0.0575)
Overweight	0.0420 (0.1007)	0.0760 (0.0547)	...	...	0.0509 (0.0957)	0.0658 (0.0549)
L.Overweight	-0.0781 (0.0519)	-0.0819** (0.0402)	...	...	-0.0625 (0.0466)	-0.0931** (0.0436)
Obese	0.0712 (0.0840)	...	0.0937 (0.0649)	...	0.0785 (0.0814)	...
L.Obese	-0.0319 (0.0582)	...	-0.0032 (0.0477)	...	-0.0085 (0.0563)	...
Severely Obese	0.0166 (0.0589)	...	...	0.0267 (0.0672)	...	0.0155 (0.0612)
L.Severely Obese	-0.0640 (0.0718)	...	...	-0.0579 (0.0647)	...	-0.0503 (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
<b>Diff-in-Hansen Tests for Exogeneity of Subsets of GMM Instruments (p-value)</b>						
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 $\geq$ 24.5		
	(1)	(2)	(3)	(4)	(5)
L.ln(wage)	0.2223*** (0.0575)	0.224*** (0.053)	0.214*** (0.059)	0.216*** (0.059)	0.220*** (0.053)
Lower BMI Var.	0.0658 (0.0549)	0.084 (0.071)	0.050 (0.148)	-0.003 (0.145)	0.018 (0.085)
L.(Lower BMI Var.)	-0.0931** (0.0436)	-0.087** (0.041)	-0.112*** (0.039)	-0.123*** (0.041)	-0.104** (0.043)
Severely Obese	0.0155 (0.0612)	...	...	0.008 (0.060)	...
L.Severely Obese	-0.0503 (0.0661)	...	...	-0.080 (0.065)	...
BMI $\geq$ 37	...	-0.138** (0.058)	...	...	-0.131** (0.056)
L.(BMI $\geq$ 37)	...	-0.069 (0.057)	...	...	-0.099* (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
<u>Diff-in-Hansen Tests for Exogeneity of Subsets of GMM Instruments (p-values)</u>					
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  $\geq$  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* (0.0412)	0.1213*** (0.0470)	...	...
Overweight	-0.1190 (0.1207)	0.0305 (0.0317)	0.0655*** (0.0166)	0.0387 (0.0271)
L.overwt	0.0549 (0.0470)	0.0115 (0.0312)	...	0.0371 (0.0269)
Severely Obese	0.0557 (0.0925)	0.0314 (0.0382)	-0.0652** (0.0266)	-0.0113 (0.0495)
L.(Severely Obese)	-0.1722*** (0.0594)	-0.0525 (0.0609)	...	-0.0813 (0.0517)
AR(1): z-statistic	-4.998	-4.746	10.57	10.59
p-value	< 0.001	< 0.001	< 0.001	< 0.001
AR(2): z-statistic	-0.568	-0.523	11.95	11.90
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 in previous tables, except for BMI variables being treated as exogenous in the second column and the OLS regressions using potential experience in place of actual, as described in the text.

**Table 5B. Preferred Specification Compared to Simpler Models**  
**White Women**

	Preferred Dynamic Specification	Dynamic w/ BMI exogenous	OLS W/out Lags	OLS W/ Lags
<i>WHO BMI Thresholds</i>				
L.ln(wage)	0.2223*** (0.0575)	0.2986*** (0.0746)	...	...
Overweight	0.0658 (0.0549)	-0.0168 (0.0342)	-0.0579*** (0.0211)	-0.0074 (0.0289)
L.overwt	-0.0931** (0.0436)	0.0151 (0.0297)	...	-0.0694** (0.0291)
Severely Obese	0.0155 (0.0612)	-0.0096 (0.0569)	0.0119 (0.0290)	0.0136 (0.0355)
L.(Severely Obese)	-0.0503 (0.0661)	-0.0234 (0.0391)	...	0.0126 (0.0368)
AR(1): z-statistic	-4.907	-4.891	10.48	10.53
p-value	< 0.001	< 0.001	< 0.001	< 0.001
AR(2): z-statistic	-0.405	-0.536	7.756	7.768
p-value	0.686	0.592	< 0.001	< 0.001
Number of Instruments	187	86	...	...
Hansen test (p-value)	0.947	0.818	...	...
<i>Alternative BMI Thresholds</i>				
L.ln(wage)	0.2197*** (0.0528)	0.3009*** (0.0730)	...	...
BMI ≥ 24.5	0.0177 (0.0852)	0.0084 (0.0277)	-0.0602*** (0.0207)	0.0069 (0.0303)
L.(BMI ≥ 24.5)	-0.1039** (0.0431)	0.0195 (0.0332)	...	-0.0904*** (0.0302)
BMI ≥ 37	-0.1305** (0.0564)	0.0615 (0.0771)	0.0269 (0.0321)	-0.0005 (0.0401)
L.(BMI ≥ 37)	-0.0990* (0.0577)	-0.0099 (0.0825)	...	0.0535 (0.0412)
AR(1): z-statistic	-5.028	-4.904	10.48	10.54
p-value	< 0.001	< 0.001	< 0.001	< 0.001
AR(2): z-statistic	-0.606	-0.604	7.761	7.773
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 in previous tables, except for BMI variables being treated as exogenous in the second column and the OLS regressions using potential experience in place of actual, as described in the text.

**Table 6. Effects of Adding Occupation to the Preferred Model**

	<u>White Men</u>		<u>White Women</u>		<u>White Women</u>	
	Preferred Specification	Occupation Added	Preferred Specification	Occupation Added	Preferred Specification	Occupation Added
L.ln(wage)	0.0720* (0.0412)	0.0721* (0.0411)	0.2223*** (0.0575)	0.2009*** (0.0529)	0.2197*** (0.0528)	0.2170*** (0.0473)
Overweight	-0.1190 (0.1207)	-0.0850 (0.1126)	0.0658 (0.0549)	0.0488 (0.0852)	...	...
L.Overweight	0.0549 (0.0470)	0.0510 (0.0463)	-0.0931** (0.0436)	-0.0604 (0.0436)	...	...
Severely Obese	0.0557 (0.0925)	0.0652 (0.1137)	0.0155 (0.0612)	0.0107 (0.0547)	...	...
L.(Severely Obese)	-0.1722*** (0.0594)	-0.2138*** (0.0790)	-0.0503 (0.0661)	-0.0622 (0.0564)	...	...
BMI ≥ 24.5	...	...	...	...	0.0177 (0.0852)	0.0081 (0.0870)
L.(BMI ≥ 24.5)	...	...	...	...	-0.1039** (0.0431)	-0.0783* (0.0455)
BMI ≥ 37	...	...	...	...	-0.1305** (0.0564)	-0.0806 (0.0520)
L.(BMI ≥ 37)	...	...	...	...	-0.0990* (0.0577)	-0.0553 (0.0540)
<u>Difference-in-Hansen Tests for Added Occupation Instruments</u>						
<i>p</i> -values	...	0.430		0.152	...	0.261

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. Occupations are grouped into five categories with service occupations being the comparison group. Second lagged levels of occupation dummies are used as instruments, producing a total of 200 instruments for men and 189 for women.

**Table 7. Effects of Health Shocks on Identification in the Preferred Models**

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