# The Impact of Obesity on Wages: the Role of Personal Interactions and Job Selection

January 3, 2018

**Abstract**

We estimate the effects of obesity on wages accounting for the workers' sorting into jobs requiring different levels of personal interactions in the workplace. Using data from the National Longitudinal Survey of Youth 1979 combined with detailed information about jobs from O*Net, we find a wage penalty for obese white women. This penalty is higher in jobs that require a high level of personal interactions. Accounting for job selection does not significantly change the estimated wage penalty.

## 1 Introduction

This paper investigates how the relationship between obesity and wages is affected by the level of personal interactions required in the workplace. A vast literature studies how obesity affects labor market outcomes.[1] In a seminal paper, Cawley (2004) finds a negative effect of obesity on wages for white women. A possible explanation of this effect is that obesity results in social stigma, affecting men and women differently

[1] See Averett and Korenman (1996), Cawley (2004), Pagan and Davila (1997), Register and Williams (1990) and references therein.

(the so-called "beauty premium").[2] To the extent that body weight affects a person's appearance, it may also affect other people's perception of his or her labor-market relevant traits.[3] If the effect of stigma on wages is more prominent for workers employed in jobs requiring frequent personal interactions, then obese workers may more likely to self-select into jobs requiring fewer interactions. We study how this self-selection affects the observed job-specific relationship between obesity and wages.

There exist at least two channels that may generate a weight penalty in jobs that require personal interactions. One possibility is that overweight individuals, being perceived as less attractive in the workplace, are discriminated against by co-workers, customers, or employers who have a taste against interacting with individuals with above normal weight (as in Becker (1957)'s theory of discrimination). In this case, jobs that require more interactions display a stronger relationship between weight and labor market outcomes. A second theory is that co-workers, customers, or employers *statistically* discriminate against overweight individuals (see Phelps (1972)). According to this theory employers hold a (possibly biased) belief that overweight workers are on average less productive, for example because they are less capable of performing the productive tasks requiring interactions. If productivity is imperfectly observed, body weight is used by employers as a proxy in order to improve workers' job allocation. Therefore, individuals carrying a different body weight, but otherwise identical, are treated differently. This (statistical) discrimination will be larger

[2] A growing literature (see Hamermesh and Biddle (1994)) documents the relationship between wages and physical appearance, which is at the root of the "beauty premium" conjecture. Bhattacharya and Bundorf (2009) offer an alternative explanation for the obesity wage penalty. Given that obese workers have higher expected medical expenditures, they also have higher health insurance costs. If these workers obtain health insurance from their employer, their employer will pass some of these costs to the employee in the form of lower wages. They find supporting evidence for this type of mechanism

[3] See Baum and Chou (2011) for an analysis of the socioeconomic causes of obesity. Another explanation from the sociological literature is that obesity has a more adverse impact on the self-esteem of white female than on that of other groups, but Averett and Korenman (1996) do not find empirical support for such conjecture.

in jobs requiring personal interactions. Both channels carry the same implication: workers in jobs that require more interactions with customers or co-workers will be more strongly affected by their appearance.

When the effects of weight on wages are economically relevant, workers may, depending on their body mass, self-select into jobs requiring different levels of personal interactions. Our contribution is to account for the endogenous selection of workers into job types by adopting a job selection model to correct for the potential selection bias occurring when selection is ignored.[4]

To this end, we merge two sources of data. The first one is the National Longitudinal Survey of the Youth 1979, a nationally representative sample of men and women aged between 14 and 22 when first interviewed in 1979. This dataset contains detailed information about the respondents, including their weight, height, employment status, occupation and wages. The second source of data is the O*Net database, which classifies occupations according to hundreds of standardized descriptors illustrating each occupations' characteristics and the worker's required skills.[5] From this information we use factor analysis to construct a variable measuring the level of personal interactions required by each occupation. We classify jobs in one of two types, depending on whether the job's level of personal interactions value is above or below the median.

The literature on the effects of obesity on wages typically regresses wages on obesity status, using different statistical methods to identify the causal effect of

[4]Previous research has noted that people with different body mass choose different jobs. Han et al. (2011), Morris (2006), while Harris (2017) examined the relationship between individuals' weight and their employment decisions over the life cycle. Our contribution is to account explicitly for the endogenous selection of people with different obesity into different types of jobs.

[5]O*Net provides for each occupation values to descriptors such as "Contact with Others" (How much does this job require the worker to be in contact with others, face-to-face, by telephone, or otherwise in order to perform it?), Face-to-Face Discussions (How often do you have to have face-to-face discussions with individuals or teams in this job?), or Work With Work Group or Team (How important is it to work with others in a group or team in this job?). The full set of descriptors used in our analysis is listed in the Appendix B.

weight (usually fixed effects or instrumental variables techniques), accounting for the possibility of reverse causality (i.e poorer individuals have less time and monetary resources to eat healthy food, exercise, etc.). We complement this approach by adopting a Roy model of self-selection.[6] In this framework, workers of different body mass may have a comparative advantage in performing different types of jobs. We assume each person chooses between two types of jobs, requiring a high or low level of personal interactions. Each job-type requires specific skills, which are distributed differently among workers of different body mass. Workers select the job that gives them the highest expected earnings. In the standard OLS regression of wages on body mass index (BMI), ignoring this selection biases the coefficient on BMI, even after controlling for job type. For example, obese individuals may disproportionally find it more advantageous to seek employment in jobs requiring fewer interactions, where obesity has less impact. Because we do not observe the wages that obese workers would obtain in jobs requiring interpersonal interactions, a standard OLS regression of wages on BMI would return a biased coefficient.

In our benchmark specifications our wage equation includes individual fixed effects to account for reverse-causality between obesity and wages, that is, the possibility that low-wages cause obesity (perhaps because poorer families have easier access to fattening food), and the possibility that unobserved correlates affect both obesity and wages.[7] To correct for the selection bias, we model job selection with an equation capturing the discrete choice between two job-types. Estimates from this equation provide information about the selection bias in the wage equation. Results from the selection equation are used, in a second stage, to unbiasedly estimate the wage equation. We implement two different approaches. In one specification, we

[6]See Roy (1951) and Heckman (1990). Seminal papers in this empirical literature are Willis and Rosen (1979), Heckman and Sedlacek (1985), and Borjas (1987).

[7]This approach is adopted in Cawley (2004).

adopt the two-step parametric method introduced by Willis and Rosen (1979) that does not include fixed effects in the selection equation. In a second specification, we use the methodology developed in Kyriazidou (1997) to include fixed effects in the selection equation.

Although the coefficients can be identified out of the functional form, to improve the identification of the bias correction, we include in the first stage a variable that we believe affects the relationship between obesity and wages only through job choice: the respondent's closest siblings' job type. The intuition for this exclusion restriction is that individuals may find it easier, regardless of their obesity, to find jobs requiring skills that are similar to the jobs of their family members, either because of direct referrals, or because family members correlate on other skills required in the workplace.

Our results confirm, using up-to-date information, the literature's result that the negative relationship between obesity and wages is significant only in some demographic groups, notably white women. However, when accounting for job selection, we see that the negative relationship between obesity and wages is mostly coming from the subset of workers in jobs needing a high level of personal interactions. In such jobs, the relationship is stronger than average, but similar to a regression that does not correct for the selection bias. In jobs that require a lower level of personal interactions, the relationship between jobs and wages is smaller and not statistically significant.

Our analysis is most related to the research in Cawley (2004), Han et al. (2009) and Shinall (2014). Cawley (2004) focuses on the effect of obesity on wages. To account for sources of bias, such as omitted variables and reverse causality, he follows different strategies: controlling for lagged values of BMI, using siblings' weight as an instrument for the respondent's weight, and, in his preferred specification, including individual fixed effects. However, he does not control for job type. Han et al. (2009) introduce the use of job characteristics. They use data from the Dictionary of Occupational Titles (DOT - a precursor of O*Net, the database we use) to obtain information about the jobs the individuals are performing. They find that the association with BMI and wages is stronger for jobs that require more interpersonal skills, but do not account for endogenous job selection. Finally, Shinall (2014) also uses job characteristics to compare the effects of obesity on wages for that require some physical skills versus jobs that require some social skills. However, in that paper the data does not offer the possibility of controlling for individual fixed effects and there is no specific accounting for selection bias in the wage-obesity estimation. Our paper's contribution relative to existing literature is to extend the analysis of the literature of obesity on wages by (i) using a broad range of information on job characteristics, and, most importantly, (ii) to explicitly model workers' job selection to verify if job sorting affects the magnitude of the effects of obesity on wages.

## 2 Data

The National Longitudinal Survey of the Youth 1979 is a nationally representative sample of the American youth. Respondents were sampled first in 1979 when they were between 14 and 22 years of age, every year until 1994, and every two years since. We use 12 years of data ranging from 1982 thru 2006.[8] This is a widely used survey in research on employment and obesity, allowing for comparability of our results with other studies. The survey contains detailed questions about employment

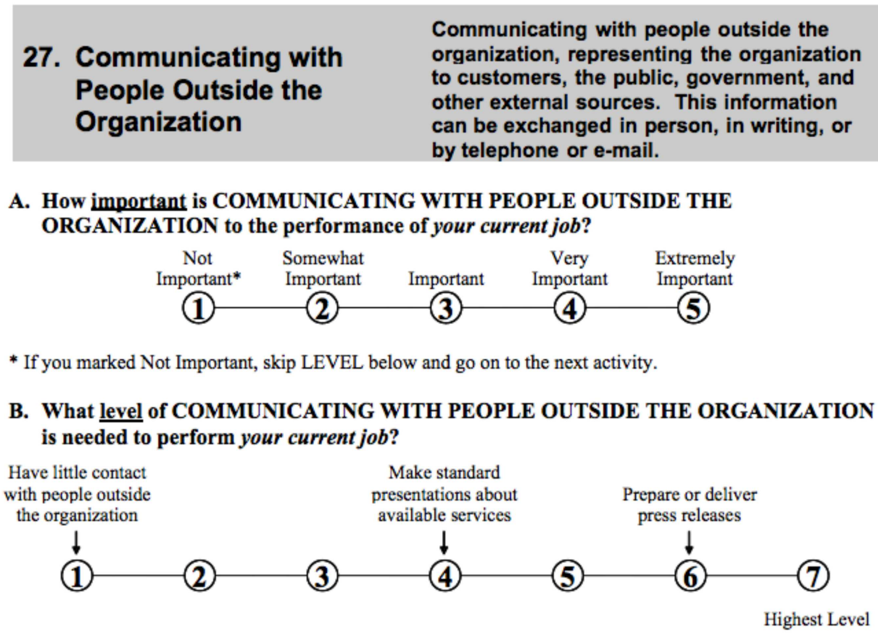
[8]The years that we end up using for our sample are: 1982, 1985, 1986, 1988, 1989, 1990, 1992, 1993, 1994, 1996, 1998 and 2006. We cannot use some of the survey years because the employment status recode we use to match the data with the O*Net database was not created in 2000-2004, or in 2008-present because the CPS section on activity in the week before the survey was not included in those rounds. In addition, the survey does not include body weight information in some of the years betweeen 1982 and 1998.

and tenure, household environment and structure, and personal health information like height and weight. For the empirical analysis we use information on gender, race, education status, marital status, number of children in the household, Armed Forces Qualification Test (AFQT), age of youngest child, highest grade achieved by mother and father, work tenure, years of experience, region dummies. We also generate dummies representing "missing" values for each of the variables to account for different types of misreporting. The estimation sample includes 41,589 person-years (19,175 person-years for women and 22,414 person-years for men) after excluding women who were pregnant at the time of interview, individuals under the age of 18, and individuals who did not have full employment or weight information. About 91 percent of our sample is employed, with on average 141 weeks of tenure at their current job.

In order to construct the main dependent outcome we obtain information on the hourly wage of each individual and then, as in Cawley (2004), we top-code the hourly wage at $500. Wages are normalized to 2010 using the Consumer Price Index for Urban Consumers. Our left hand side variable is the log of real hourly wages. Our estimation sample shows that the average wage for men is $20.23 while the average wage for women is $16.25. When focusing on the difference between normal weight category and obese category, we estimate a wage differential of $-2.83 for women and a $1.44 for men, i.e. obese men earn - on average - higher hourly wages than normal weight men.

In order to calculate body mass index, we pool the responses for all years that recorded self-reported weight. Height was assumed to be equal to the height recorded in 1985, when respondents were between 20 and 27 years of age. We adopt the standard practice in this literature to correct weight and height for self-reporting error using the procedure proposed by Lee and Sepanski (1995), exploiting the informa-

Figure 1: Sample of Question about Importance and Level in O*Net



**27. Communicating with People Outside the Organization**

**Communicating with people outside the organization, representing the organization to customers, the public, government, and other external sources. This information can be exchanged in person, in writing, or by telephone or e-mail.**

**A. How important is COMMUNICATING WITH PEOPLE OUTSIDE THE ORGANIZATION to the performance of *your current job*?**

| Not Important* | Somewhat Important | Important | Very Important | Extremely Important |
|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 |

* If you marked Not Important, skip LEVEL below and go on to the next activity.

**B. What level of COMMUNICATING WITH PEOPLE OUTSIDE THE ORGANIZATION is needed to perform *your current job*?**

| Have little contact with people outside the organization | | | Make standard presentations about available services | | Prepare or deliver press releases | Highest Level |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |

tion on the relationship between true and reported height and weight collected in the National Health and Nutrition Survey. In our estimation sample the average BMI is around 25.92 for women and 26.65 for males. Almost 22 percent of women and about 21 percent of males in our sample falls in the obese or higher category.

We identified the respondents' job in every year in which they are employed, as well as the associated hours worked and the hourly wage received at each job. If individuals are working multiple jobs, they report one of the jobs as their main job and we use that information as their main source of employment. The NLSY offers census occupation codes for all jobs. We use these codes to merge the NLSY data with data from from the 2010 revision of the Occupational Information Network (O*Net) database, which contains detailed information about job characteristics.

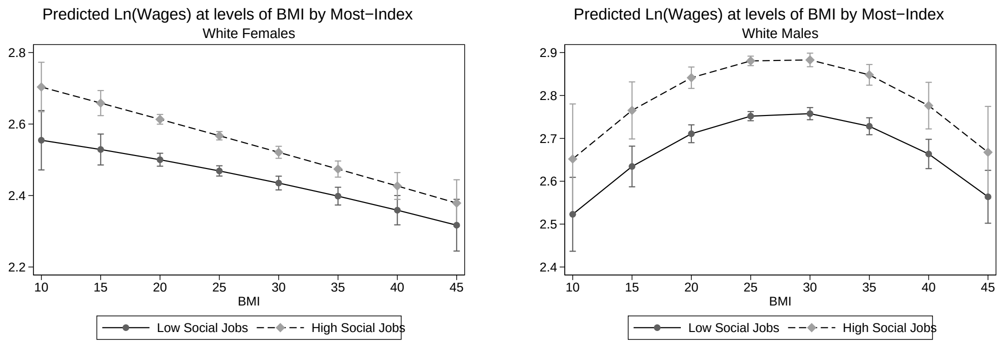
For each job, the O*Net offers a description of tasks, tools and technologies, knowledge, skills, abilities, work activities, work context and education required. For each piece of information, the database reports two numerical values, "importance" and "level". The "importance" value is a number ranging from 1 to 5 representing how how often the skill is used; the "level" is a number ranging from 1 to 7 representing the expertise in skill needed to perform the job. For example, the skill "writing" may be equally important for secretaries and journalists, but a professional writer may need a higher level of writing skills. The database contains a total of 277 job descriptors. Figure 1 reports, as an example, one of the questions which is relevant to infer the importance and level of personal interactions required to perform the respondent's job.

We use information from the categories "skills" and "abilities" in O*Net to calculate how important personal interactions are to perform any given job. We chose 17 work activities and job skills that, to our judgement, are most important for discerning the important personal interactions. These questions assess how relevant such interactions are to perform a job, to communicate with others (supervisors, co-workers, or customers), to resolve conflicts or coordinate other individuals, to speak, negotiate, coordinate others, etc. The full list of job descriptors we selected can be found in Appendix B.

We used the "level" values for each task and skill; to reduce the dimensionality of the information we used factor analysis to extract a single variable measuring the degree of personal interactions required by each job.[9] We use the median of this standardized score for the entire sample as a threshold to categorize jobs in two categories: those requiring high and low levels of personal interaction. In our estimation sample we find that 71% of women are occupied in jobs requiring a high level of personal interactions, whereas only 60% of males are employed in these

[9]We replicated the analysis using the "importance" indicators, which lead to similar results. In our main specifications we use the "level" measure because it displays more variation than the "importance" measure.

Figure 2: Wages and BMI for different groups and job types



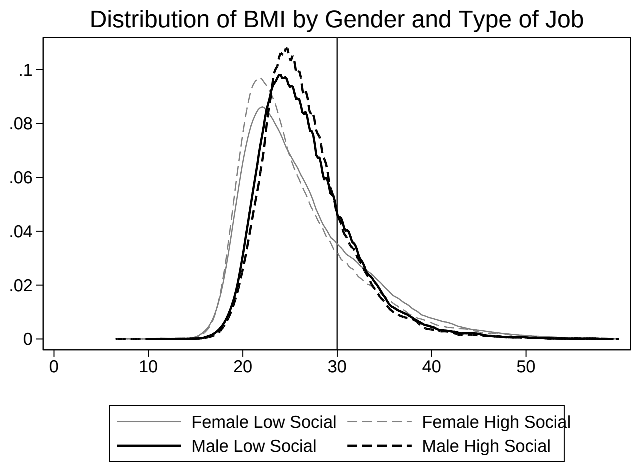

job. In Figure 2, we plot the predicted logarithm wage across the different levels of BMI for jobs requiring a high level of interactions versus jobs requiring low levels of personal interactions for two groups, white women and white males.[10]

These figures illustrate several facts. First, jobs requiring interactions usually pay higher wages than other jobs. Second, the relationship between BMI, wages, and job-type is different between males and women. The relationship between wages and BMI is close to linear for women, and decreasing in BMI; the same relationship for males is non-monotonic. Finally, the wage-gap between jobs with high and low levels of personal interactions for white women is decreasing in BMI, suggesting that there is a higher penalty of BMI on wages for women in jobs requiring interactions, presumably because weight and appearance is more important in such jobs. The same penalty is not as evident for males' wages as for women' wages.

Figure 3 displays the distribution of BMI by type of job and gender. The BMI distribution appears to display a smaller variance in jobs requiring more interactions, suggesting the possibility of job selection which we will explore in the empirical analysis.

[10]The other race-gender groups can be found in Appendix C

Figure 3:




## 3 Empirical framework

We now turn to describing our empirical strategy to account for the selection into the different types of jobs. We assume that individuals can be employed in two types of jobs $j$, requiring ($j = 1$) or not requiring ($j = 2$) a high level of personal interactions. Let $w_{itj}$ be the log wage of individual $i$ at time $t$ employed in job $j \in \{1, 2\}$, which we assume to depend on a set of covariates $X_{it}$, including her or his obesity at time $t$, $\text{BMI}_{it}$, and, depending on the adopted specification, individual and time effects to account for the possibility of reverse causality from wages to BMI:

$$w_{itj} = \alpha_j X_{it} + \zeta_i + \epsilon_{jit} \tag{1}$$

Let $Z_{it}$ denote a set of observed variables that influence the job choice of individual $i$ at time $t$ without affecting wages directly, and let $J_{it}$ be a latent variable determining the job-type choice of the individual. We assume the following job selection model:

$$J_{it} = \beta Z_{it} + \eta_i - \epsilon_{3it} \tag{2}$$

with job choice $j_{it} = 1$ if $J_{it} \geq 0$, and $j_{it} = 2$ otherwise. Hence,

$$\Pr(j_{it} = 1) = \Pr(\epsilon_{3it} \leq \beta Z_{it} + \eta_i) \tag{3}$$

Because individuals selects into different jobs according to their characteristics, wages are not observed from a random sample of the population and an OLS regression of the wage equation with fixed effects delivers biased coefficients because:

$$\begin{aligned} E(w_{it1}|j_{it} = 1) = \alpha_1 X_{it} + \zeta_i + E(\epsilon_{1it}|j_{it} = 1) &= \alpha_1 X_{it} + \zeta_i + E(\epsilon_{1it}|\epsilon_{3it} \leq \beta Z_{it}) \\ E(w_{it2}|j_{it} = 2) = \alpha_2 X_{it} + \zeta_i + E(\epsilon_{2it}|j_{it} = 2) &= \alpha_2 X_{it} + \zeta_i + E(\epsilon_{2it}|\epsilon_{3it} > \beta Z_{it}) \end{aligned}$$

Procedures to account for the selection bias in linear models have been developed since the seminal paper by Heckman (1979). We estimate the model (1-2) using two different specifications that account for selection and, for comparison, we estimate a specification that does not account for selection.

### 3.1 Specification without selection (OLS-FE)

First we include a standard specification which does not account for job selection. We include fixed effects in the wage equation as in the preferred specification of Cawley (2004), but separate the sample according to the type of job (high or low personal interactions). Essentially, we estimate (1) for the sample of individual working in jobs with high personal interactions and low level of personal interactions separately.

### 3.2 Linear parametric specification with fixed effects in the wage equation (WR)

In our first specification that accounts for selection we use the standard approach to selection, and we include individual fixed effects in the wage equation (the second stage). First, we assume joint normality of the error term vector $[\epsilon_1, \epsilon_2, \epsilon_3]$, with zero means and variance-covariance matrix $\Sigma = [\sigma_{ij}]$. This normality assumption allows the computation of an analytical expression for the bias:

$$E(w_{it1}|j_{it} = 1) = \alpha_1 X_{it} + \frac{\sigma_{13}}{\sigma_{33}} \frac{\phi\left(\frac{\beta Z_{it}}{\sigma_{33}}\right)}{\Phi\left(\frac{\beta Z_{it}}{\sigma_{33}}\right)} \tag{4}$$

$$E(w_{it1}|j_{it} = 2) = \alpha_2 X_{it} + \frac{\sigma_{23}}{\sigma_{33}} \frac{\phi\left(\frac{\beta Z_{it}}{\sigma_{33}}\right)}{\left(1-\Phi\left(\frac{\beta Z_{it}}{\sigma_{33}}\right)\right)} \tag{5}$$

where $\phi$ is the PDF and $\Phi$ is the CDF of a standardized normal distribution. These equations can be estimated with the two-step procedure adopted in Willis and Rosen (1979): first, estimate the probit model (3) and use the estimates to compute the predicted values of the inverse Mills ratios[11]:

$$\hat{\lambda}_{1it} = \frac{\phi\left(\frac{\beta \hat{Z}_{it}}{\sigma_{33}}\right)}{\Phi\left(\frac{\beta \hat{Z}_{it}}{\sigma_{33}}\right)} \tag{6}$$

$$\hat{\lambda}_{2it} = \frac{\phi\left(\frac{\beta \hat{Z}_{it}}{\sigma_{33}}\right)}{1-\Phi\left(\frac{\beta \hat{Z}_{it}}{\sigma_{33}}\right)} \tag{7}$$

Next, estimate the wage equations (1) by Ordinary Least Squares including $\hat{\lambda}_{1it}$ and $\hat{\lambda}_{2it}$ to correct for the selection bias. The second stage then becomes:

$$\ln(w_{it1}) = \alpha_1 X_{it} + \zeta_i + \gamma_1 \hat{\lambda}_{1it} + \epsilon_{1it}$$

$$\ln(w_{it2}) = \alpha_2 X_{it} + \zeta_i + \gamma_2 \hat{\lambda}_{2it} + \epsilon_{2it}$$

The procedure provides consistent estimates of the parameter vector $\alpha$.

In this approach identification of model parameters is guaranteed by the nonlin-

[11] See, e.g., Hogg and Craig (1995)

earity of the Mills ratios. However, non-parametric identification is usually preferable if a variable is available that affects the selection equation without affecting the wage equation directly. To this end, we include in $Z_{it}$ the closest sibling's job type. We motivate this choice with the assumption that, because jobs are often found through family mentoring and networking, a sibling's job type has a more direct effect on the other sibling's job selection than on her wages. This exclusion restriction will improve identification of the parameter of interest, the coefficient on BMI.[12]

All specifications for the wage equation include as regressors the following variables: individual fixed effects, categorical BMI dummies, age, number of children, age of youngest child, education, work tenure and experience, region dummies, marital status, mother's education and experience

### 3.3 Specification with fixed effects in both the wage and the selection equations (K)

Our third and preferred specification adds one more complication, adding fixed effects in *both the selection and the wage equation*. The standard procedure in estimating fixed effects models with panel data is to compute time-differences of the equation of interest, which eliminates the individual fixed effects $\zeta_i$. However, if the selection equation also contains fixed effects, sample selectivity creates nonlinearities in the wage equation that cannot be differenced out.

We therefore adopt the method proposed in Kyriazidou (1997) to account for individual unobserved heterogeneous effects in both the selection and the wage equations. The procedure follows a two-step approach in the spirit of Heckman (1979), in which the unknown coefficients of the selection equation (2) consistently estimated

[12] In order to identify the "closest" sibling of each individual we create a measure of distance between each individual and all of their siblings. This measure is the sum of the squared difference between the siblings' age, race and gender. The "closest" possible sibling is one that has the same age, race and gender.

in the first step are used to estimated the equation of interest 1 in the second step. The wage equation is estimated with the usual procedure of taking time differences of the observed variables as in standard linear panel data models, which eliminates the fixed effects $\zeta_i$. To account for sample selectivity, each observation is assigned a weight computed in the first step. The procedure assumes that, for an individual choosing job $j$ in two consecutive periods, the selection effect remains the same if the variables that determine the job choice do not change over time, in which case the selection effect is completely eliminated by the time differencing in the second stage. A Kernel weight is therefore computed as a function of the magnitude of the estimated differences $|w_{i,t+1}\hat{\beta} - w_{i,t}\hat{\beta}|$ so that a larger difference corresponds to a smaller weight, where $\hat{\beta}$ can is consistently estimated using a logistic model. In other words, for a given individual, if we observe moves from different job types and no change in BMI, then this observation will not receive too much weight in the wage equation, as opposed to individuals who experience changes in both jobs and BMI.

In this specification we include the same covariates included in the previous specifications, exdluding the parental variables that, not varying over time, are not identified when fixed effects are included in the first stage.

# 4 Results

In order to allow for appropriate comparisons, we use the same sample in all specifications.[13] Common sample size and average wages for all specifications are reported in Table 1, and descriptive statistics in Appendix A. As standard practice in the

[13]An OLS regression that includes all observations available (including those without siblings' information) produces somewhat different results, with bigger effects of obesity on wages. The presence of siblings obviously affects family size, which in standard models of fertility is correlated with human capital of the parents (and their children, if there is intergenerational transmission of skills), and human capital may differentially affect the relationship between BMI and wages (Sarlio-Lähteenkorva et al. (2004)).

| | High personal interactions | | | Low personal interactions | | |
|---|---|---|---|---|---|---|
| | $N \times T$ | $N$ | $\overline{ln(w)}$ | $N \times T$ | $N$ | $\overline{ln(w)}$ |
| Whites | | | | | | |
| Women | 6,998 | 1,450 | 2.67 | 4,260 | 1,291 | 2.34 |
| Men | 6,514 | 1,358 | 2.96 | 6,145 | 1,475 | 2.59 |
| Blacks | | | | | | |
| Women | 2,384 | 584 | 2.52 | 2,560 | 667 | 2.26 |
| Men | 1,699 | 523 | 2.77 | 4,025 | 797 | 2.45 |
| Hispanics | | | | | | |
| Women | 1,763 | 389 | 2.61 | 1,210 | 341 | 2.31 |
| Men | 1,625 | 397 | 2.86 | 2,406 | 491 | 2.58 |

Table 1: Sample sizes and mean log wages

literature, we look at the effect on wages of categorical values of BMI: individuals are defined underweight if their BMI is below 18.5, overweight if their BMI is between 25 and 30, and obese if their BMI is greater than 30. The main results are reported in tables 2 (Whites), 3 (Blacks), and 4 (Hispanics). In each table, columns 1, 2, and 3 reports the results for specifications 1, 2, and 3, respectively, as described in the previous section. The tables display the effects of the categorical BMI levels (normal BMI is the omitted category) on wages. The full set of coefficient estimates, including results from the first stage regressions for specifications (1) and (2) are reported in the external appendix. Standard errors are clustered at the individual level. The top panel of each table reports the estimates for women and the bottom panel for men. In each panel we report the results for individuals in jobs that require high levels of personal interactions and for individuals in jobs that require a low level of personal interactions.

Taking a bird-eye view by looking for traction on the coefficients of the categorical BMI variables across different groups and specifications, we note that abnormal BMI has consistent statistically and economically significant effects only on white women working in jobs requiring a high level of personal interactions. This group suffers a wage penalty from being obese between 10 and 11 percent, depending on the specification. The wage penalty associated from being simply overweight is smaller in size, but significant and consistent across specifications.

The effect of abnormal BMIs on black and hispanic women in jobs with a high level of personal interactions is in almost all specifications smaller in magnitude and noisier, and rarely statistically significant. For black women who select into jobs with high levels of personal interaction, we find in our preferred specification a strong positive effect of being underweight. Since this result does not seem to be robust across specifications, we take it to provide only suggestive evidence that there is a premium for underweight black women among jobs that have high levels of personal interactions. Overweight status is also associated with higher wages (relative to the missing category, normal BMI). However this effect becomes less significant once individual fixed effects are included in the selection equation.

The effects of abnormal BMI in women working in job with low levels of personal interactions are generally smaller in magnitude for all racial groups. In the rare cases when they are statistically significant, the results are not robust to different specifications. For example, obese black women receive a wage premium of about 10 percent, in these types of jobs, barely significant at the 5 percent level according to specifications (1) and (2), but this result is reversed, but noisy in our preferred specification (column 3). Looking at the results from the Hispanic women sample, we don't see any statistically significant effects of the categorical weight variables for either type of job.

The effect of abnormal BMI on men is generally small and positive, but rarely significant, and not robust to changing specifications. For example, we find statistically significant positive effects of being overweight in white males working in jobs requiring high personal interactions. However, this effect is not robust to the inclusion on individual fixed effects in the selection equation.

Comparing our results to existing literature, we confirm the previous findings of a wage penalty on for obese white women, but not for other groups. However, we find that these effects are stronger, and statistically significant, only in jobs requiring a high level personal interactions. These patterns are consistent with the results from Figure 2 and suggest that white women's wage penalty for obesity is driven, in part, by social stigma rather than productivity. This is consistent with the hypothesis that appearance matters more in these types of jobs.

Finally, we note that accounting for selection, with or without including fixed effects in the selection equation, has only modest effects on the estimated coefficients.

Table 2: Effects of BMI on log wages for whites

| **Women** | (1) OLS-FE | (2) WR | (3) K |
|---|---|---|---|
| *High personal interactions* | | | |
| Underweight | -0.090* | -0.107* | -0.042 |
| | (0.042) | (0.043) | (0.049) |
| Overweight | -0.063* | -0.070* | -0.075* |
| | (0.029) | (0.029) | (0.037) |
| Obese | -0.109* | -0.109* | -0.097* |
| | (0.047) | (0.047) | (0.045) |
| Mills Ratio | | -0.359 | |
| | | (0.185) | |
| *Low personal interactions* | | | |
| Under weight | -0.066 | -0.066 | -0.04 |
| | (0.050) | (0.050) | (0.054) |
| Overweight | 0.003 | 0.003 | -0.019 |
| | (0.026) | (0.026) | (0.044) |
| Obese | -0.067 | -0.067 | -0.094 |
| | (0.050) | (0.050) | (0.081) |
| Mills Ratio | | -0.000 | |
| | | (0.000) | |

| **Men** | (1) OLS-FE | (2) WR | (3) K |
|---|---|---|---|
| *High personal interactions* | | | |
| Underweight | 0.030 | 0.032 | -0.051 |
| | (0.077) | (0.078) | (0.225) |
| Overweight | 0.051* | 0.051* | -0.025 |
| | (0.025) | (0.025) | (0.052) |
| Obese | 0.049 | 0.045 | -0.046 |
| | (0.040) | (0.042) | (0.077) |
| Mills Ratio | | 0.119 | |
| | | (0.215) | |
| *Low personal interactions* | | | |
| Underweight | -0.098 | -0.098 | 0.027 |
| | (0.123) | (0.123) | (0.101) |
| Overweight | 0.043 | 0.043 | 0.104 |
| | (0.024) | (0.024) | (0.052) |
| Obese | 0.017 | 0.017 | 0.04 |
| | (0.035) | (0.035) | (0.063) |
| Mills Ratio | | 0.000 | |
| | | (0.000) | |

Table 3: Effects of BMI on log wages for blacks

| **Women** | (1) OLS-FE | (2) WR | (3) K |
|---|---|---|---|
| *High personal interactions* | | | |
| Underweight | -0.141 | -0.147 | 0.205* |
| | (0.093) | (0.094) | (0.072) |
| Overweight | 0.085* | 0.074 | -0.009 |
| | (0.041) | (0.045) | (0.075) |
| Obese | 0.095 | 0.086 | -0.074 |
| | (0.059) | (0.059) | (0.099) |
| Mills Ratio | | -0.178 | |
| | | (0.337) | |
| *Low personal interactions* | | | |
| Under weight | -0.040 | -0.040 | -0.19 |
| | (0.068) | (0.068) | (0.102) |
| Overweight | 0.056 | 0.056 | 0.001 |
| | (0.031) | (0.031) | (0.049) |
| Obese | 0.098* | 0.098* | -0.094 |
| | (0.049) | (0.049) | (0.077) |
| Mills Ratio | | 0.000 | |
| | | (0.000) | |

| **Men** | (1) OLS-FE | (2) WR | (3) K |
|---|---|---|---|
| *High personal interactions* | | | |
| Underweight | 0.028 | 0.030 | -0.024 |
| | (0.205) | (0.200) | (0.000) |
| Overweight | 0.025 | 0.041 | 0.069 |
| | (0.050) | (0.047) | (0.033) |
| Obese | 0.039 | 0.054 | 0.125* |
| | (0.062) | (0.062) | (0.059) |
| Mills Ratio | | 0.310 | |
| | | (0.263) | |
| *Low personal interactions* | | | |
| Underweight | 0.057 | 0.057 | -0.132 |
| | (0.063) | (0.063) | (0.183) |
| Overweight | 0.029 | 0.029 | 0.089 |
| | (0.025) | (0.025) | (0.044) |
| Obese | 0.062 | 0.062 | 0.119 |
| | (0.041) | (0.041) | (0.066) |
| Mills Ratio | | 0.000 | |
| | | (0.000) | |

Table 4: Effects of BMI on log wages for hispanics

| **Women** | (1) OLS-FE | (2) WR | (3) K |
|---|---|---|---|
| *High personal interactions* | | | |
| Underweight | -0.129 | -0.317* | -0.126 |
| | (0.084) | (0.123) | (0.091) |
| Overweight | 0.012 | 0.057 | 0.014 |
| | (0.054) | (0.058) | (0.129) |
| Obese | 0.015 | 0.124 | -0.034 |
| | (0.088) | (0.099) | (0.131) |
| Mills Ratio | | -1.714* | |
| | | (0.686) | |
| *Low personal interactions* | | | |
| Under weight | -0.084 | -0.086 | 0.118 |
| | (0.070) | (0.069) | (0.101) |
| Overweight | 0.030 | 0.031 | -0.085 |
| | (0.037) | (0.037) | (0.053) |
| Obese | 0.008 | 0.008 | -0.071 |
| | (0.063) | (0.063) | (0.078) |
| Mills Ratio | | 0.000 | |
| | | (0.000) | |

| **Men** | (1) OLS-FE | (2) WR | (3) K |
|---|---|---|---|
| *High personal interactions* | | | |
| Underweight | 0.769 | 0.764 | 1.112 |
| | (0.406) | (0.407) | (0.814) |
| Overweight | 0.012 | 0.015 | -0.338* |
| | (0.058) | (0.067) | (0.117) |
| Obese | 0.050 | 0.055 | -0.199 |
| | (0.108) | (0.113) | (0.122) |
| Mills Ratio | | -0.070 | |
| | | (0.470) | |
| *Low personal interactions* | | | |
| Underweight | -0.115 | -0.115 | -0.116 |
| | (0.074) | (0.074) | (0.047) |
| Overweight | 0.014 | 0.013 | -0.031 |
| | (0.026) | (0.026) | (0.037) |
| Obese | 0.093* | 0.093* | -0.1 |
| | (0.044) | (0.044) | (0.054) |
| Mills Ratio | | -0.000 | |
| | | (0.000) | |

# 5 Discussion

Our main results are consistent with the findings in Cawley (2004) and Han et al. (2009) (see Table 5 for a comparison of their main results with our specification (2)).[14] Both of these papers find significant negative effects of obesity on wages for white women, and the magnitude of these estimates is within the confidence interval of our unadjusted estimates. Similarly, these papers do not find strong evidence for the effects of obesity in males or black and hispanic minorities. This reinforces the evidence for an obesity wage penalty occurring for white women, but not other race-gender groups. Recall that while our data source is the same as in Cawley (2004) and Han et al. (2009), differently from them, our sample only includes individuals with siblings.

We find that most of the negative effect of obesity for white women is due to women working in jobs requiring a high level of personal interactions. However, we find suggestive evidence that weight causes women to sort into different types of jobs and that this effect could be biasing the standard OLS downwards. These results help validate previous findings on the estimation of BMI on wages since most of those studies do not correct for selection into these types of job.

Focusing on white women, our results highlight that body mass is a relatively more relevant factor in jobs requiring a high level of personal interactions, suggesting that employers may consider looks in addition to productivty when it comes to jobs that require high levels of personal interation.

This methodology has some limitations and one should interpret a causal interpretation of the coefficients with caution. First, even though we are accounting for

[14] There are several differences across the different specifications: Cawley (2004) and Han et al. (2009) do not include year 2006. Han et al. (2009) wage equations are conditioned on people who are employed. Cawley (2004) and Han et al. (2009) use a blue collar dummy and several state and local economic characteristics.

Table 5: Comparison of our results with Cawley (2004) and Han et al. (2009)

| | | | Our results | | | |
|---|---|---|---|---|---|---|
| | Cawley (2004) | Han et al. (2009) | High OLS | Low OLS | High (K) Adjusted | Low (K) Adjusted |
| White women | -0.087 | -0.075 | -0.109* | -0.067 | -0.097* | -0.094 |
| | (0.015) | (0.021) | (0.047) | (0.050) | (0.045) | (0.081) |
| Black women | 0.002 | -0.049 | 0.095 | 0.098* | -0.074 | -0.097 |
| | (0.017) | (0.025) | (0.059) | (0.049) | (0.099) | (0.077) |
| Hisp. women | -0.020 | 0.032 | 0.015 | 0.008 | -0.034 | -0.071 |
| | (0.024) | (0.035) | (0.088) | (0.063) | (0.131) | (0.078) |
| White males | 0.013 | -0.001 | 0.049 | 0.017 | 0.046 | 0.04 |
| | (0.015) | (0.017) | (0.040) | (0.035) | (0.077) | (0.063) |
| Black males | 0.031 | 0.014 | 0.039 | 0.062 | 0.125* | 0.119 |
| | (0.019) | (0.027) | (0.062) | (0.041) | (0.059) | (0.066) |
| Hisp. males | 0.023 | 0.008 | 0.050 | 0.093* | -0.199 | -0.10 |
| | (0.025) | (0.032) | (0.108) | (0.044) | (0.122) | (0.054) |

Coefficients on the category for "Obese" in each paper. Robust standard errors are in parentheses. The specification in Cawley (2004) and Han et al. (2009) include individual fixed effects.

selection, there could be some unobservable time-varying components across individuals that would be biasing the effects of obesity on wages. Second, the variable we exclude from the wage equation in the selection models could be correlated with wages through some unobserved variable. Finally, we do not model the possible changing intertemporal preferences of individuals that may vary with age and family status, and could be affecting their employment decision. Despite these limitations, our results extend the literature of obesity and wages by providing evidence that the impact of BMI on wages depends on job types but not on sorting.

# A Appendix: descriptive statistics

Table 6: Summary Statistics of NLSY 1987-2006

| | Women | | | Men | | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hispanic | White | Black | Hispanic | Total |
| ***Employment Variables*** | | | | | | | |
| Employed | 0.96 | 0.92 | 0.94 | 0.96 | 0.91 | 0.93 | 0.95 |
| Hourly Wage in Dollars in 2010 $ | 16.77 | 13.42 | 14.97 | 20.98 | 15.62 | 19.37 | 18.39 |
| Ln(Wage) in 2010 $ | 2.58 | 2.40 | 2.50 | 2.82 | 2.55 | 2.71 | 2.68 |
| High personal interactions job | 0.64 | 0.48 | 0.59 | 0.54 | 0.30 | 0.42 | 0.56 |
| Tenure (weeks) | 186.3 | 184.4 | 171.7 | 213.6 | 173.0 | 188.2 | 197.0 |
| Currently in School | 0.12 | 0.08 | 0.09 | 0.09 | 0.06 | 0.07 | 0.10 |
| Experience (1000s hours) | 17.47 | 16.34 | 16.87 | 22.42 | 19.49 | 21.03 | 19.81 |
| ***Body Mass Index*** | | | | | | | |
| BMI Adjusted | 24.63 | 28.20 | 26.35 | 26.22 | 26.32 | 26.95 | 25.76 |
| Underweight (BMI $<$ 18.5) | 0.04 | 0.02 | 0.02 | 0.00 | 0.01 | 0.01 | 0.02 |
| Overweight (25 $<$ BMI $<$ 30) | 0.21 | 0.28 | 0.30 | 0.38 | 0.35 | 0.41 | 0.31 |
| Obese (BMI: $>$30) | 0.14 | 0.33 | 0.22 | 0.17 | 0.19 | 0.20 | 0.17 |
| ***Demographics*** | | | | | | | |
| Age | 29.68 | 30.41 | 30.12 | 29.97 | 30.19 | 29.81 | 29.90 |
| Highest Grade | 13.67 | 13.17 | 12.53 | 13.35 | 12.50 | 12.14 | 13.35 |
| Mother Highest Grade | 11.78 | 10.33 | 7.99 | 11.54 | 10.10 | 7.45 | 11.24 |
| Father Highest Grade | 11.91 | 8.01 | 7.20 | 11.87 | 7.75 | 7.14 | 11.12 |
| Number of Children | 0.97 | 1.44 | 1.40 | 0.89 | 1.26 | 1.21 | 1.00 |
| Age of Youngest Child | 2.43 | 4.06 | 3.60 | 1.76 | 1.77 | 1.86 | 2.20 |
| Never Married | 0.29 | 0.47 | 0.28 | 0.38 | 0.53 | 0.41 | 0.36 |
| Married | 0.57 | 0.33 | 0.55 | 0.52 | 0.33 | 0.47 | 0.52 |
| Separated | 0.02 | 0.10 | 0.07 | 0.02 | 0.06 | 0.04 | 0.03 |
| Divorced | 0.11 | 0.09 | 0.10 | 0.08 | 0.07 | 0.07 | 0.09 |
| Widowed | 0.00 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |

Table 7: Summary Statistics of NLSY 1987-2006 for women

| | Underweight | Normal | Overweight | Obese | Total |
|---|---|---|---|---|---|
| BMI Adjusted | 17.63 | 21.84 | 27.14 | 35.41 | 25.17 |
| ***Employment Variables*** | | | | | |
| Employed | 0.92 | 0.96 | 0.96 | 0.95 | 0.96 |
| Hourly Wage in Dollars in 2010 $ | 15.70 | 17.15 | 15.56 | 14.26 | 16.25 |
| Ln(Wage) in 2010 $ | 2.48 | 2.59 | 2.53 | 2.45 | 2.55 |
| High personal interactions job | 0.62 | 0.63 | 0.62 | 0.57 | 0.62 |
| Tenure (weeks) | 122.2 | 167.7 | 197.5 | 242.5 | 185.3 |
| Currently in School | 0.16 | 0.13 | 0.09 | 0.06 | 0.11 |
| Experience (1000s hours) | 11.40 | 15.79 | 18.90 | 21.54 | 17.30 |
| ***Demographics*** | | | | | |
| Age | 26.86 | 28.80 | 30.71 | 32.58 | 29.80 |
| Highest Grade | 13.24 | 13.74 | 13.48 | 13.10 | 13.55 |
| Mother Highest Grade | 11.78 | 11.74 | 11.13 | 10.57 | 11.40 |
| Father Highest Grade | 11.35 | 11.67 | 10.85 | 9.90 | 11.17 |
| Number of Children | 0.88 | 0.90 | 1.24 | 1.39 | 1.06 |
| Age of Youngest Child | 2.07 | 2.20 | 3.20 | 3.84 | 2.70 |
| Never Married | 0.36 | 0.33 | 0.28 | 0.32 | 0.32 |
| Married | 0.42 | 0.53 | 0.59 | 0.52 | 0.54 |
| Separated | 0.05 | 0.03 | 0.04 | 0.05 | 0.04 |
| Divorced | 0.17 | 0.11 | 0.08 | 0.11 | 0.11 |
| Widowed | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

Table 8: Summary Statistics of NLSY 1987-2006 for Males

| | Underweight | Normal | Overweight | Obese | Total |
|---|---|---|---|---|---|
| BMI Adjusted | 17.50 | 22.77 | 27.14 | 33.69 | 26.27 |
| ***Employment Variables*** | | | | | |
| Employed | 0.84 | 0.94 | 0.96 | 0.96 | 0.95 |
| Hourly Wage in Dollars in 2010 $ | 15.85 | 19.24 | 21.25 | 20.68 | 20.23 |
| Ln(Wage) in 2010 $ | 2.35 | 2.72 | 2.86 | 2.81 | 2.78 |
| High personal interactions job | 0.36 | 0.49 | 0.53 | 0.49 | 0.50 |
| Tenure (weeks) | 86.9 | 165.2 | 230.4 | 267.4 | 207.1 |
| Currently in School | 0.11 | 0.12 | 0.06 | 0.05 | 0.08 |
| Experience (1000s hours) | 10.75 | 17.56 | 24.17 | 28.90 | 21.98 |
| ***Demographics*** | | | | | |
| Age | 25.45 | 28.08 | 30.98 | 32.90 | 29.99 |
| Highest Grade | 12.00 | 13.23 | 13.24 | 12.90 | 13.17 |
| Mother Highest Grade | 10.11 | 11.15 | 11.12 | 10.99 | 11.11 |
| Father Highest Grade | 9.86 | 11.23 | 11.13 | 10.59 | 11.07 |
| Number of Children | 0.45 | 0.76 | 1.07 | 1.24 | 0.96 |
| Age of Youngest Child | 0.45 | 1.16 | 2.01 | 2.81 | 1.76 |
| Never Married | 0.66 | 0.50 | 0.33 | 0.30 | 0.40 |
| Married | 0.24 | 0.40 | 0.57 | 0.59 | 0.50 |
| Separated | 0.02 | 0.03 | 0.02 | 0.02 | 0.02 |
| Divorced | 0.07 | 0.07 | 0.08 | 0.09 | 0.08 |
| Widowed | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 |

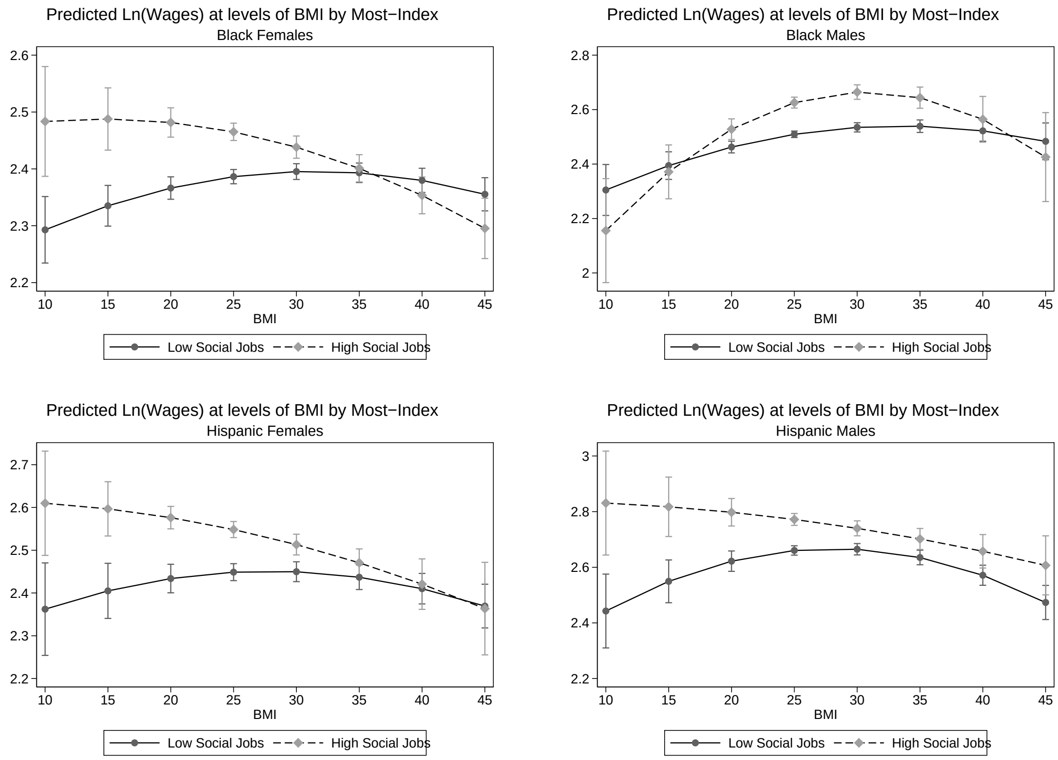
## B Appendix: job descriptors

We report below the set of abilities, skills and work activities we used to construct the index that encompasses the level of personal interaction required on the job.

- Work Activities:
  - Communicating with Supervisors, Peers, or Subordinates
  - Communicating with Persons Outside Organization
  - Establishing and Maintaining Interpersonal Relationships
  - Assisting and Caring for Others
  - Selling or Influencing Others
  - Resolving Conflicts and Negotiating with Others
  - Coordinating the Work and Activities of Others
- Skills:
  - Speaking
  - Social Perceptiveness
  - Coordination
  - Persuasion
  - Negotiation
  - Instructing
  - Service Orientation

# C Relationship between BMI and wages, other groups

Figure 4: Relationship between BMI and wages across race and gender groups




# References

**Averett, Susan and Saners Korenman**, "The economic reality of The Beauty Myth," *The Journal of Human Resources*, 1996, *31* (2), 304–30.

**Baum, Charles L and Shin-Yi Chou**, "The Socio-economic causes of obesity," Technical Report, National Bureau of Economic Research 2011.

**Becker, Gary S.**, *The Economics of Discrimination*, University of Chicago Press, Chicago, 1957.

**Bhattacharya, Jay and M Kate Bundorf**, "The incidence of the healthcare costs of obesity," *Journal of health economics*, 2009, *28* (3), 649–658.

**Borjas, George J.**, "Self-Selection and the Earnings of immigrants," *American Economic Review*, September 1987, *77* (4), 531–553.

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

**Hamermesh, Daniel and Jeff E. Biddle**, "Beauty and the Labor Market," *American Economic Review*, 1994, *84* (5), 1174–94.

**Han, Euna, Edward C Norton, and Lisa M Powell**, "Direct and indirect effects of body weight on adult wages," *Economics & Human Biology*, 2011, *9* (4), 381–392.

**_ , Edward Norton, and Sally Stearns**, "Weight and Wages: Fat Versus Lean Paychecks," *Health Economics*, 2009, (18), 535–548.

**Harris, Matthew**, "The Impact of Body Weight on Occupational Mobility and Career Development," *Working Paper*, 2017.

**Heckman, James J.**, "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, *47* (1), 153–161.

_ , "The Empirical Content of the Roy Model," *Econometrica*, September 1990, *58* (5), 1121–1149.

**Heckman, James J and Guilherme Sedlacek**, "Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market," *The Journal of Political Economy*, December 1985, *93* (6), 1077–1125.

**Hogg, Robert V. and Allen T. Craig**, *Introduction to Mathematical Statistics (5th edition)*, Upper Saddle River, New Jersey: Prentice Hall, 1995.

**Kyriazidou, Ekaterini**, "Estimation of a panel data sample selection model," *Econometrica*, 1997, pp. 1335–1364.

**Lee, Lung-Fei and Jungsywan H. Sepanski**, "Estimation of linear and non-linear errors-in-variables models using validation data," *Journal of the American Statistical Association*, 1995, *90* (429), 130–140.

**Morris, Stephen**, "Body mass index and occupational attainment," *Journal of Health Economics*, 2006, *25* (2), 347–364.

**Pagan, Jose A and Alberto Davila**, "Obesity, occupational attainment, and earnings: Consequences of obesity," *Social Science Quarterly*, 1997, *78* (3), 756–770.

**Phelps, Edmund S**, "The Statistical Theory of Racism and Sexism," *The American Economic Review*, 1972, *62* (4), 659–661.

**Register, Charles A and Donald R Williams**, "Wage effects of obesity among young workers," *Social Science Quarterly*, 1990, *71* (1), 130–141.

**Roy, Andrew Donald**, "Some thoughts on the distribution of earnings," *Oxford economic papers*, 1951, *3* (2), 135–146.

**Sarlio-Lähteenkorva, Sirpa, Karri Silventoinen, and Eero Lahelma**, "Relative Weight and Income at Different Levels of Socioeconomic Status," *American Journal of Public Health*, 2004, *94* (3), 468–472.

**Shinall, Jennifer**, "Why Obese Workers Earn Less: Occupational Sorting and its implications for the legal system," *Social Science Research Network*, 2014.

**Willis, Robert J. and Sherwin Rosen**, "Education and Self-Selection," *Journal of Political Economy*, October 1979, *87* (5), S7–S36.