Causal effects of BMI on wage

Sankar Mukhopadhyay[1] and Joe Crouse

Department of Economics

University of Nevada Reno

June 2014

This is a preliminary draft prepared for 13th IZA/SOLE conference.

Please do not cite without permission.

Abstract: We provide evidence that negative effects of BMI on wage are different in different parts of wage distribution. We use instrumental variable quantile regression. We use a novel instrument which is the age at menarche of a sibling as the instrument for BMI of a female respondent. In particular we show that marginal effect of BMI on wage increases with BMI till the 80th percentile as one would normally expect. However, at the 90th percentile the marginal effect of BMI on wage decreases with an increase in BMI.

[1] Corresponding author, Department of Economics, University of Nevada Reno, Reno, NV, 89557. Email: sankarm@unr.edu

Introduction

Obesity has become a salient global health concern in recent years due to its profound negative economic and social consequences. According to the Centers for Disease Control and Prevention (2012), the percentage of obese people in the United States rose from 19.4% in 1997 to an alarming 27.7% in 2011. Lakdawalla & Philipson (2009) state that more Americans are obese than smoke, use illegal drugs, or suffer from ailments unrelated to obesity.

The objective of this paper is to estimate the effect of obesity on earnings for women. Several studies (Register & Williams, 1990; Averett & Korenman, 1996; Pagan & Davilla, 1997; Behrman and Rosenzweig, 2001 ; Cawley, 2004; Atella et al., 2008; Johar & Katayama, 2012) that have found an inverse relationship between earnings and body weight for females. In this paper we revisit this issue. However, there are several important innovations in our paper compared to the literature.

1) We use a new instrument for BMI. We use age at menarche of a female sibling as an instrument for BMI of a respondent. Several studies have shown that age at menarche and adult BMI are strongly correlated (Trikudanathan et.al., 2013; Pearce and Leon 2005). Pearce and Leon (2005) show that age at menarche has an effect on adult BMI even after controlling for childhood BMI. They conclude that age at menarche is a marker for sexual maturation which in turn affects adult BMI. Age at menarche therefore provides us with exogenous variation in the BMI of a sibling. Furthermore, BMI of a sibling has been widely used as an instrument in this literature (Cawley, 2004; Johar & Katayama, 2012). Therefore it is plausible that age of menarche of a sibling would be correlated with the BMI of a female. The advantage is that if sibling BMI is a valid instrument (there is some

evidence in the literature it is), age at menarche of a sibling will be valid. However, the converse is not necessarily true.

2) This literature thus far has pursued linear IV models with the exception of Kline and Tobias (2010) and Gregory and Rhum (2011). However, it is rather difficult to theoretically justify why a movement from underweight to normal-weight category will come with a wage penalty. Both the studies that allow for nonlinear relationship between BMI and wage find that the relationship in indeed nonlinear. Furthermore the existence of non-linear relationship may help in exploring the reasons behind the negative relationship between wage and BMI. For example Gregory and Rhum (2011) find that the decline in women's wage happens even in the "healthy" BMI range. This result leads them to conclude that the obesity penalty is not being driven by extra healthcare expenditure on obese workers as claimed by Bhattacharya & Bundorf (2009). We allow for potential non linearity as well. We use a control function approach to deal with the endogeneity, and therefore we can allow for nonlinearity in the endogenous variable.

3) Just as the effect of BMI on wage may be different on different parts of BMI distribution (i.e. the nonlinearity issue discussed above) there is no reason to think, ex ante, that the effect of BMI is same at all wage levels. There are only two studies in the literature, to the best of our knowledge that utilize quantile regression techniques in estimating the effect of obesity on earnings. Atella et al. (2008) use quantile regression but do not address the issue of potential endogeneity of BMI in their analysis. Johar & Katayama (2012) use an IV quantile regression technique with the same-sex siblings' body mass as an instrument (following Cawley, 2004). However, neither paper allow for

nonlinearity in the effects of BMI on wage. Johar & Katayama (2012) find that body mass and wages are negatively correlated at all points in the wage distribution for females and the strength of this relationship is greater at higher-wage levels. Therefore we use instrumental variable quantile regression (Lee, 2007) to determine whether the effect of BMI is different in different parts of the wage distribution.

Methods

Here we are briefly presenting the instrumental variable quantile regression method as developed by Lee (2007) for application purposes. This method uses control function approach to address endogeneity and is different from instrumental variable quantile regression developed by Chernozhukov and Hansen (2004; 2006). Further details are available in Lee (2007). We want to estimate the following quantile regression equation

$$lnW = B\beta_1(\tau) + B^2\beta_2(\tau) + X\gamma(\tau) + U$$

Where B is BMI and B^2 is squared BMI of an individual. We include both BMI and BMI squared in the wage equation because we expect that the effect on wage of an increase in BMI in the “healthy” range may be different from an increase in BMI in the “unhealthy” range. B is potentially endogenous and therefore so is B^2. X is a vector of control variables and U is the structural error term. $\beta(\tau)$ and $\gamma(\tau)$ are structural parameters with $0 < \tau < 1$. Given that we have endogenous regressors in our structural equation we have to use instrumental variable regression. In other words we need one or more variables that is correlated with satisfies the exclusion restriction. We plan to use the age at menarche of a female sibling as the instrument for BMI. Therefore linear instrumental variable quantile regression methods cannot be used. Let A be the instrument

$$B = X\pi(\alpha) + A\theta(\alpha) + V$$

Where V is reduced for error and $0 < \alpha < 1$. Lee (2007) page 3 outlines the assumptions required for this model. First one is that conditional on X and A, τ-th quantile of the structural error (U) is a function of reduced form error (V), and the second one requires quantile independence of the reduced form error (V) . To estimate the model we used a modifiled version of Stata command “cqiv” developed by y Chernozhukov et al., (2012). The “cqiv” command was modified by the authors to obtain bootstrapped standard errors after adjusting for clustering.

Data

The data used in this study are from The National Longitudinal Survey of Youth (NLSY97) which was designed to represent the entire population of American youth. All sample members were between 12 to 16 years of age in 1997 when the first annual interview was conducted. The interviews have been conducted every year since then. This paper utilizes the available data from 1997 to 2011. In this paper we only report the results for white females. Results from black and Hispanic females are in the appendix.

Respondents are asked to report their race or ethnicity at the baseline of the NLSY97. As is conventional in this literature, we categorize the three groups as Black, Hispanic, and White. Our sample consists of 27.6% black individuals, 21.1% Hispanic individuals, and 51.3% white individuals. In constructing our dataset, males are excluded because of the nature of our instrumental variables approach. Women that are pregnant in a given year are excluded for that yearly observation since women face a situation of unusual weight fluctuation during and after pregnancy that cannot be attributed to influencing wages in the same way that obesity does.

The NLSY97 recorded the self-reported heights and weights of respondents in each year from 1997 to 2011. The body mass index is calculated as $\frac{weight\ in\ pounds}{height\ in\ inches^2} X\ 703.0696$[2].

Women who are pregnant in a given year are dropped from the sample in that particular year. We include general intelligence measured by the Armed Services Vocational Aptitude Battery (ASVAB) test, and highest grade completed to control for differences in human capital attainment among individuals. In order to control for employment characteristics we control for years of actual work experience and current job tenure, current school enrollment, whether the respondent's job is part-time[3], and whether they have employer provided health insurance. Age, marital status, number of children, and region of residence are also included as regressors in our model.

Our dependent variable in the study is the hourly wage earned by the respondent at her primary job. NLSY97 calculates this variable each year. Table 1 presents the descriptive statistics for our key variables. Our sample for white women consists of 3511 person year observations from 460 women. The average wage in our sample is $16.83 per hour (in 2011 prices). 26.1% of respondents were high school graduates, 28.4% had some college education but were not college graduates, and 25.3% were college graduates. The average age in our sample is 22.7 years. Average work experience is 6.7 years, and

[2] The clinical classifications are: underweight (BMI < 18.5), healthy weight (18.5 < BMI < 25), overweight (BMI > 25), and obese (BMI > 30).

[3] Part-time status is defined as less than 20 hours of work per week at the primary job.

average job tenure is 1.8 years. 20% are married and average number of children is 0.4. Given the average age these numbers are not surprising.

The average BMI in our sample is 25.43. Average age at menarche for a female sibling is 12.7 years in our sample which is similar to 12.6 reported in Chumlea et al (2003) for White females in the U.S.

Results

We start by reporting OLS results. As shown in Table 2 OLS estimates suggest that one unit increase in BMI is associated with 1.2% decline in hourly wage for white females. Cawley (2003) reported a decline of 1.0% with that one unit increase in BMI. Next we implement linear instrumental variable estimation using sibling BMI -- the same instrument as Cawley (2003). The first stage F stat for excluded instrument is 53.7 suggesting that the instrument is strong. Our IV estimates suggest that one unit increase in BMI leads to 2.9% decline in hourly wage. IV estimates from Cawley (2003) suggest 1.7% decline in hourly wage. Thus while our estimates are somewhat larger than the estimates obtained by Cawley's (2003) using data from NLSY79, they are largely consistent with Cawley (2003).

Next we implement the IV estimates with sibling's age at menarche. This instrument has not been used in this literature before. As we argued above this instrument is more likely to be valid that sibling's BMI. Column 3 of Table 2 presents the IV estimates with this instrument. The first stage F stat for excluded instrument is 7.0 suggesting that there may be a weak instrument problem and the estimates should be treated with caution. However, all the weak-instrument robustness checks (Anderson-

Rubin Wald test statistic has a p-value of 0.007, and Stock-Wright LM S statistic has a p-value of 0.014) suggest that the coefficient of BMI in the wage equation is different from zero. The IV estimate in this case suggests that an one unit increase in BMI leads to 5.9% decline in hourly wage. This is almost five times as large as the OLS estimate.

Next we focus on quantile regression instrumental variable estimates. As discussed above we allow for non-linear effects of BMI on wage and we allow these effects to be different across the wage distribution. Table 3 presents the results of QIV estimates with sibling's age at menarche as the instrument[4].

1) For the median wage worker a one unit increase in BMI in the "healthy" range (BMI between 20 and 25) is associated with 0.4% decline in wage. However, a one unit increase in BMI in the "obese" range (BMI between 30 and 35) is associated with a larger 1.1% decline in wage. In general we observe that one unit increase in BMI causes a larger and larger percentage decline in wage as BMI goes up. The estimated effects for the median wage worker are therefore not very different from the OLS estimates.
2) Up to the 80th percentile of the wage distribution we observe that the negative effect of BMI increases with BMI. However, at the 80th percentile effect is flat and at the 90th percentile of wage distribution the negative effect of BMI is highest in the "healthy" range and then declines with an increase in BMI.

At this point it is worthwhile to compare our results to the two other papers that have allowed nonlinear effects of BMI on wage: Kline and Tobias (2010) and Gregory and

[4] Standard errors of these estimates are calculated using bootstrapping. Bootstrapping these estimators are time consuming. They are not available yet. They will be calculated soon.

Rhum (2011). Both papers report that for white females wage starts to decline while their BMI is well below 25 (i.e. still in the "healthy" range). We find the same result. However, both of those papers suggest that the negative effect of BMI on wage is highest when BMI is in mid 20s (i.e. BMI between 23 and 28). However, our results suggest that those results are most likely an artifact of mean based estimator. We show that the effect of BMI on wage is very different on different parts of wage distribution with the effect of BMI increasing with BMI till the 80th percentile, flat at the 80th percentile, and declining at the 90th percentile. A mean based regression is most likely aggregating these patterns to produce the results reported in Kline and Tobias (2010) and Gregory and Rhum (2011).

We noted earlier in the results section that sibling's age at menarche is relatively weak. The other instrument which is sibling BMI is strong and has been widely used in this literature. Next we present the results of instrumental variable quantile regression with sibling BMI as the instrument. Panel A of Table 4 presents the results of QIV estimates with sibling as the instrument. The QIV estimates using BMI sibling's age at menarche for this sample are presented in panel B on Table 4. Estimates from panel A show that one unit increase in BMI causes a larger and larger percentage decline in wage as BMI increases up to the 80th percentile. This pattern is reversed after the 80th percentile of wage distribution where the negative effect of BMI is highest in the "healthy" range and then declines with an increase in BMI. Panel B shows the of instrumental variable quantile regression with sibling age at menarche as the instrument but now we are using the same sample as in panel A. The results are very similar. In fact for the median wage worker even the point estimates are exactly the same.

The result that for most of the wage distribution the negative effect of BMI on wage increases with BMI is what one would have anticipated ex-ante. This is consistent with all three explanations discussed in the literature. First, marginal productivity may decline as a person becomes overweight or obese. Next, employer and/or consumer discrimination may cause employers to be reluctant to hire obese people (Morris, 2007) or put them in jobs that do not require interaction with customers (Han, Norton, and Stearns 2009, Han, Norton and Powell, 2011). Finally, obese individuals may experience lower wages because employers spend relatively more on the healthcare of obese employees compared to non-obese employees (Bhattacharya & Bundorf, 2009). In this case, compensation is hypothesized to be the same: wages reduce to equilibrate the increase in benefits of obese individuals. Next we restrict our attention to only those observations where the worker does not have employer provided health insurance. But even with this restriction we observe the same pattern as described above. The results are very similar irrespective of which instrument is used. Since these workers do not have employer provided health insurance an increased BMI should not lead to lower wages if higher healthcare cost was the reason. Therefore these results suggest that higher healthcare cost on obese workers may not be the only reason behind obesity penalty.

Conclusion

We provide evidence that negative effects of BMI on wage are different in different parts of wage distribution. In particular we show that marginal effect of BMI on wage increases with BMI till the 80th percentile as one would normally expect. However, at the 90th percentile the marginal effect of BMI on wage decreases with an increase in

BMI. This is somewhat puzzling and perhaps suggests that discrimination may be an important factor at the top of the wage distribution.

REFERENCES

Atella, V., Pace, N. & D. Vuri (2008). Are employers discriminating with respect to weight? European evidence using quantile regression. *Economics and Human Biology* 6: 305-329.

Averett, S. & S. Korenman (1996). The conomic Reality of the Beauty Myth. *Journal of Human Resources* 31(2): 304-30.

Baum, C. & W. Ford (2004). The wage effects of obesity: a longitudinal study. *Health Economics* 13: 885-899.

Bhattacharya, J. & M. Bundorf (2009). The incidence of the healthcare costs of obesity. *Journal of Health Economics* 28: 649-658.

Cawley, J. (2004). The Impact of Obesity on Wages. *The Journal of Human Resources* 39(2): 451-474.

Chumlea, William Cameron, Christine M. Schubert, Alex F. Roche, Howard E. Kulin, Peter A. Lee, John H. Himes, and Shumei S. Sun. "Age at menarche and racial comparisons in US girls." *Pediatrics* 111, no. 1 (2003): 110-113.

Goel, R. K. (2006). Obesity: An Economic and Financial Perspective. *Journal of Economics and Finance* 30(3): 317-324.

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

Johar, M.. & H. Katayama (2012). Quantile regression analysis of body mass and wages. *Health Economics* 21: 597-611.

Lakdawalla, D., & Philipson, T. (2009). The Growth of Obesity and Technology Change. *Economics & Human Biology* 7(3): 283-293.

Majumder, M. A. (2013). Does obesity matter for wages? Evidence from the United States. *Economics Papers* 32(2): 200-217.

Pagan, J. & A. Davila (1997). Obesity, Occupational Attainment, and Earnings. *Social Science Quarterly* 8(3): 756-70.

Register, C. & D. Williams (1980). Wage Effects of Obesity among Young Workers. *Social Science Quarterly* 71(1):130-41.

Smith, G.D., Sterne, J.A.C., Fraser, A., Tynelius, P., Lawlor, D.A., Rasmussen, F., 2009. The association between BMI and mortality using offspring BMI as an indicator of own BMI: large intergenerational mortality study. BMJ 339 (December), b5043.

Trikudanathan, S & et.al. (2013). Association of female reproductive factors with body composition: the Framingham Heart Study. *Journal of Clinical Endocrinology & Metabolism* 98(1): 236-44.

Wada, R. & E. Tekin (2010). Body composition and wages. *Economics and Human Biology* 8: 242-254.

Table 1: Summary statistic

	Mean/percentage	SD
Real wage (2011 $)	17.05	4.91
Age	22.84	3.82
Tenure	1.85	2.08
Work experience	6.69	3.61
High school graduate	0.26	0.44
Some college	0.28	0.45
College graduate	0.26	0.44
Married	0.21	0.41
Enrolled	0.37	0.48
Number of Children	0.41	0.80
Northeast	0.19	0.39
Northcentral	0.30	0.46
South	0.33	0.47
Insurance	0.39	0.49
BMI	24.52	5.47
Sibling BMI	24.66	5.58
Age at menarche	12.76	1.31
Sibling age at menarche	12.73	1.30
	4069	

	OLS		IV (sibling BMI)		IV (sibling age at menarche)	
	Coefficient	sd	Coefficient	sd	Coefficient	sd
BMI	-0.013	-5.85	-0.028	-3.83	-0.059	-2.25
age	0.019	0.38	0.041	0.72	0.086	1.24
age2	0.000	-0.08	-0.001	-0.42	-0.001	-0.93
yrten	0.057	4.45	0.059	4.42	0.056	4.05
yrten2	-0.004	-3.08	-0.005	-3.25	-0.005	-3.24
exp	0.009	0.55	0.016	0.88	0.011	0.56
exp2	0.000	0.27	0.000	0.14	0.001	0.58
ehs	0.065	2.08	0.069	1.97	0.060	1.44
esomecoll	0.153	4.06	0.134	3.19	0.130	2.54
ecoll	0.373	7.54	0.335	5.87	0.286	3.73
married	0.045	1.37	0.054	1.48	0.057	1.5
enrollment_	-0.058	-2.18	-0.043	-1.56	-0.073	-2.37
children_	-0.035	-1.71	-0.032	-1.43	-0.017	-0.59
northeast	-0.001	-0.03	0.004	0.09	0.042	0.73
northcentral	-0.091	-2.35	-0.084	-1.98	-0.067	-1.34
south	-0.063	-1.6	-0.062	-1.45	-0.026	-0.52
insurance_	0.202	6.87	0.225	7.15	0.239	6.4
_cons	1.329	2.43	1.384	2.32	1.565	2.56
# obs	4069		3609		4069	

Table 3: QIV estimates: effect of BMI on log wage

	Wage quantile								
BMI range	10	20	30	40	50	60	70	80	90
15-20	0.7	0.4	0.6	0.2	-0.1	-0.3	-0.4	-1.1	-3.4
20-25	0.2	-0.1	0.1	-0.2	-0.4	-0.6	-0.7	-1.2	-2.8
25-30	-0.4	-0.5	-0.4	-0.6	-0.8	-0.9	-1.1	-1.3	-2.2
30-35	-0.9	-0.9	-0.9	-1.0	-1.1	-1.2	-1.5	-1.3	-1.6
35-40	-1.4	-1.3	-1.4	-1.4	-1.4	-1.5	-1.9	-1.4	-1.0

Table 4

Panel A								
	Wage quantile							
10	20	30	40	50	60	70	80	90
-0.2	0.0	0.5	-0.3	-0.6	-0.6	-0.7	-1.6	-3.5
-0.3	-0.3	0.0	-0.5	-0.7	-0.8	-0.9	-1.4	-2.9
-0.4	-0.6	-0.5	-0.6	-0.8	-1.0	-1.1	-1.2	-2.3
-0.5	-0.9	-1.0	-0.8	-0.9	-1.1	-1.3	-1.0	-1.6
-0.6	-1.2	-1.5	-1.0	-1.0	-1.3	-1.5	-0.9	-1.0
Panel B								
-0.3	-0.2	0.2	-0.3	-0.6	-0.8	-1.0	-1.5	-3.7
-0.4	-0.4	-0.2	-0.5	-0.7	-0.9	-1.1	-1.3	-3.0
-0.4	-0.6	-0.5	-0.6	-0.8	-1.0	-1.1	-1.2	-2.3
-0.5	-0.8	-0.8	-0.8	-0.9	-1.1	-1.2	-1.0	-1.6
-0.5	-1.0	-1.1	-1.0	-1.0	-1.2	-1.3	-0.9	-0.9