The Effect of Depression on Obesity:
an instrumental variable approach

Marco Goretti

`marco.goretti@unil.ch`

Abstract

This article studies the effect of depression on obesity through the use of the spouse's depression as an instrument to correct the simultaneity bias. A strongly significant effect was found on females while a barely significant effect was found on males, excepted in the morbidly obese category where the gender does not play a role anymore.

Having smoked in the past or being currently smoking greatly reduced the probability of being obese and the gender did not play a role on this effect.

UNIL | Université de Lausanne
HEC Lausanne

`https://github.com/mgoretti/HEC-microeconometrics`

1 Introduction

Awareness and discussion about obesity have been steeply rising in the last 40 years, alongside the affected population. This subject is particularly relevant in the US which makes it the perfect place for studies.

Depression has also been the subject of recent discussions, on one hand regarding its role and prevalence in society to encourage better acceptance and also, on the other hand, about the recent results regarding its link with neuroinflammatory diseases (Setiawan et al. [2015]).

Several studies found that depression and obesity were linked, such as Markowitz et al. [2008], which proposed a bidirectional model and suggested treating them as a comorbid condition for health care purposes, Luppino et al. [2010], which confirmed a reciprocal link through a meta-analysis and Dave et al. [2011] which found a seven percentage points increase in the probability of being overweight or obese in women with current or past depression diagnosis (no significant effect in men) and linked this causality to an increase of the economic burden of depression by about 10% (9.7 billion $).

Regarding the different prevalence of depression between sexes, Silverstein [2002] found that somatic symptoms of depression were much higher among women. Obesity could be classified in the later category.

This paper aims at confirming the previously cited effect of depression on obesity through the use of the depression of the spouse as an instrument for the endogenous depression of the subject.

2 Data

2.1 Source

The study has been conducted on the Health and Retirement Study[1] database which consists of a panel of approximatively 20'000 US residents over the age of 50 that are surveyed every 2 years (currently 11 waves of which the last 10 are used).

2.2 Raw data transformation

The wide-form panel was transformed into a long form and those observations were considered as separate (clustered by individual and controlled for the wave number to take into account the increase of obesity over time).

All observations with missing variables were dropped (resulting in 100'600 observations left).

---

[1]`http://hrsonline.isr.umich.edu/`

Dummies for weight problems (BMI < 18.5|BMI > 25), overweight (BMI > 25), obese (BMI > 30) and morbidly obese (BMI > 40) were created.

2.3 Depression Score

The intensity of depression is measured using the CESD (Center for Epidemiologic Studies Depression) scale which asks 6 negative and 2 positive questions and sums the scores (1 if yes to a negative question or no to a positive question) yielding a score that goes from 0 (least depressed) to 8 (most depressed).

2.4 Choice of Instrument

Given the endogenous relationship between obesity and depression, an instrument for depression had to be found. The chosen one was the depression of the spouse which is highly correlated with the individual depression and, as it can be seen in table 1, less correlated with obesity. We argue that most of the correlation is through the subject's depression (the spouse is depressed because the partner is depressed, not because he/she is fat). Thus leading to an exogenous instrument.

	obese	CESD self	CESD spouse
obese	1		
CESD self	0.0755	1	
CESD spouse	0.0430	0.2245	1

Table 1: Correlations between obesity and depressions on the CESD scale

Another instrument was needed for the interaction effect between gender and depression and was constructed by multiplying the spouse depression score by the gender dummy.

3 Descriptive Analysis

The original dataset contained 37'317 persons, after transforming it to long format, 410'487 observations where observed which were reduced to 100'600 after removing missing informations (most of the persons did not participate through the whole study.

Figure 2 shows the positive correlation between level of depression and obesity while figure 3 shows the symmetry of the behaviour (people at both sides of body weight problems have a higher level of depression).

Figure 4 shows that depression seems stationary with respect to the age while figure 5 shows a decreasing trend in the bmi as people grow older.

It is also worth noticing the upward trend in weight as times passes (more recent wave) from figure 7 and the absence of obvious trend in the evolution of depression from figure 6.

3.1 Pooled OLS

The pooled OLS method was chosen, disregarding the possible advantages of using a better multi-dimensional analysis method, such as as using fixed effect panel data.

This decision was motived by 2 reasons:

- Many relevant unobserved variables, such as diet, local cuisine, stress and work situation were not going to stay constant through the period as so the observation for one person should not be considered as strongly linked as fixed effects does.

- BMI changes slowly, thus making a great part of observations constant which removes a great number of subjects that never changed their obesity nor depression situation.

4 Results

4.1 Regression Model

To answer the research question of this paper, the effect of depression on obesity, the following basis of regression model was used:

$$\text{Obese} = \beta_0 + \beta_1 \cdot \text{CESD} + \beta_2 \cdot \text{Male} + \beta_3 \cdot \text{Male} \cdot \text{CESD} + \beta_4 \cdot \text{Smoked} + \beta_5 \cdot \text{Male} \cdot \text{Smoked} + \delta \cdot \text{wave} + \xi \cdot \chi + \epsilon$$

Where wave is a vector of dummies telling in which wave the observation is and ξ is composed of the controls education level, age and smoking situation of the spouse.

We then computed this model using 3 regressions types:

- A linear regression

- A linear regression using the spouse's CESD score as instrument for CESD score and the spouse's interaction for the interaction effect. (gender times CESD score)

- The above regression using an instrumental probit instead of the linear regression

Auxiliary regressions were also performed using weight problems, overweight and morbidly obese as the explained variable instead of obese.

3

4.2 Main results

Table 4 clearly shows a strongly positive influence of the depression score on the probability of becoming obese (almost 5% at the mean). What is also worth noting is that females are severely more affected by this (being a male more than halves the coefficient, but still keeps it significantly different from zero) even though males are more likely to be obese in the baseline situation (6.5% in the non-depressed case).

It is also worth noting that older people are less likely to be obese but this result suffers from endogeneity (being obese decreases the life expectancy) and also from the age distribution of the database (only takes the older part of the population into account) and since it was not the focus of the study, trying to improve the identification of this effect through a better functional form did not make sense. The age of the spouse does not have any significant effect.

Having smoked strongly decreased the probability of being obese (-15%) and the effect did not seem to vary through sexes, while the smoking situation of the spouse did not have any significant effect.

Regarding the controls, as it can be seen in table 6 and 7, the wave (thus the year of the observation) has a positive influence on the probability of being obese and a higher education reduces that probability.

4.3 Strong Instrument

The chosen instruments are very strong as was observed with the Cragg-Donald Wald F statistic of over 2000 in all the regressions.

It is interesting to observe that the use of instruments actually increases the amplitude of the coefficient compared to the regression with the endogenous regressor. As it can be seen in table 8, the simplest regression roughly keeps the same coefficients as the complete one. The simpler regression allows to compute the direction of the bias: reducing the bias by using an instrument increased the coefficient, which implies that the bias was negative, thus the regressor (depression score) is negatively correlated with the errors which suggests a negative effect of obesity on depression using (with α_1 the effect of depression on obesity and α_2 the effect of obesity on depression):

$$\text{Cov}(x, u) = \frac{\alpha_2}{1 - \alpha_1 \alpha_2} \sigma_\epsilon^2$$

conditional on $\alpha_1 \alpha_2 < 1$ which seems plausible since α_1 is very small.

4.4 Auxiliary results

The coefficient of interest behave similarly in all 4 cases and it can also be noted that the ivreg and the marginal effects of the iv profit are very similar suggesting that we are in the middle of the distribution (in the linear part). The only situation where there is a visible difference between the

4

coefficient when comparing ivreg with ivprobit (however not significant) is in the morbidly obese case where we only have a small percentage of the population in one category thus putting the regression far from the linear part at the center.

The behaviour of the coefficients of interest can be seen in figure 1 and 8 :
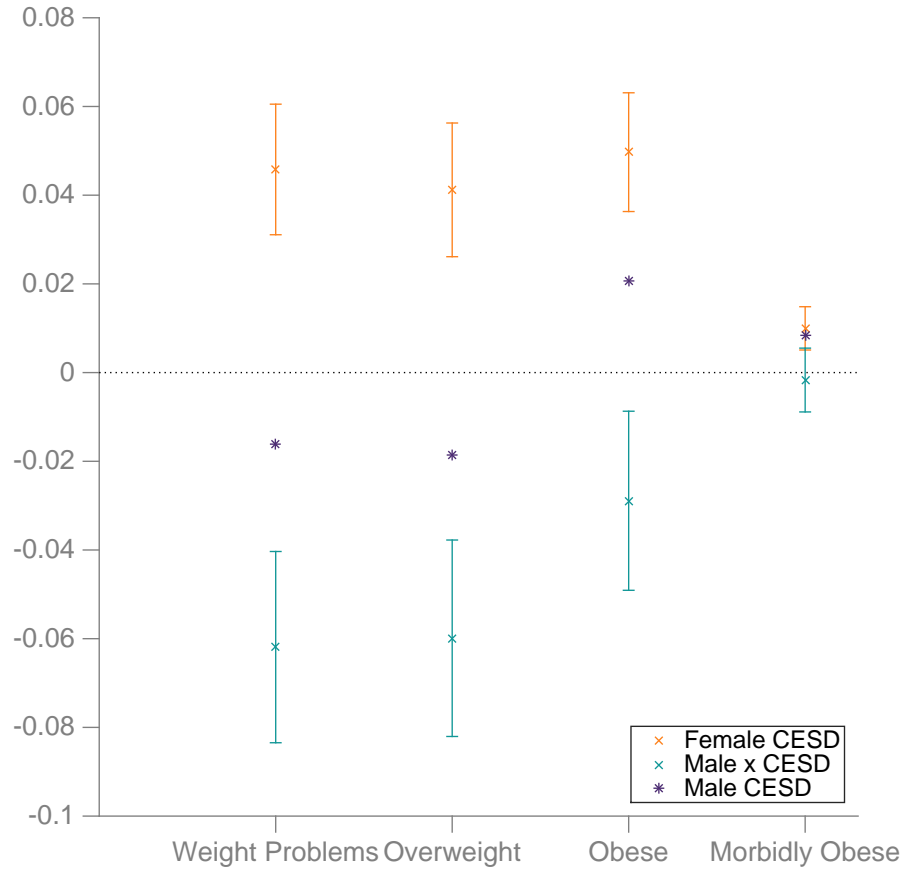
Figure 1: Visual representation of the obtained coefficients for CESD and both sexes

As it can be seen, the effect of depression on being overweight or obese in female is very strong while being way lower (even changing sign in the broader categories) for males. However it is interesting to see that this difference disappears when looking at the probability of being morbidly obese where the sex does not play a role anymore.

Finally, the gender does not change the effect of smoking for any dependent variable.

5 Conclusion

In this paper we showed that depression has a positive effect on obesity. When also considering the interaction effect with the gender, the cumulative effect was still significant in both sexes but was way stronger in women for any weight problem besides morbid obesity, on which depression has an equal positive effect for both sexes. These findings are in line with Markowitz et al. [2008], Luppino et al. [2010] and Dave et al. [2011] while also confirming the higher manifestation of somatic symptoms of depression in women found by Silverstein [2002].

The novelty of the presented results consists in the use of the spouse's depression as an instrument for the subject's depression to correct the simultaneity bias and also finding a negative effect of depression on being overweight in males.

It must be noted that the obtained results are restricted to the special observed population which is composed of US residents with an over representation of older persons and also only applies to married individuals because of the needed instrument.

References

Dhaval M. Dave, Jennifer Tennant, and Gregory J. Colman. Isolating the effect of major depression on obesity: Role of selection bias. *NBER Working Paper*, 2011.

Floriana S. Luppino, Leonore M. de Wit, Paul F. Bouvy, Theo Stijnen, Pim Cuijpers, Brenda W. J. H. Penninx, and MD Frans G. Zitman. Overweight, obesity, and depression: A systematic review and meta-analysis of longitudinal studies. *Arch Gen Psychiatry*, 2010.

Sarah Markowitz, Michael A. Friedman, and Shawn M. Arent. Understanding the relation between obesity and depression: Causal mechanisms and implications for treatment. *CLINICAL PSYCHOLOGY: SCIENCE AND PRACTICE*, 15, 2008.

Elaine Setiawan, Alan A. Wilson, Romina Mizrahi, Pablo M. Rusjan, Laura Miler, Grazyna Rajkowska, Ivonne Suridjan, James L. Kennedy, P. Vivien Rekkas, Sylvain Houle, and Jeffrey H. Meyer. Role of translocator protein density, a marker of neuroinflammation, in the brain during major depressive episodes. *JAMA Psychiatry*, 2015.

Brett Silverstein. Gender differences in the prevalence of somatic versus pure depression: A replication. *The American Journa of Psychiatry*, 2002.

6 Appendix

6.1 Github

All code used for this project can be found at

https://github.com/mgoretti/HEC-microeconometrics

6.2 IV Probit settings

Due to the complexity of the model and number of observations, the iterating algorithm computing the maximum likelihood was having trouble converging (and taking way too much time to do so). I identified the problem in the criteria for the convergency of the method. It was using an absolute difference instead of a relative one (and a very high number of max iterations). In a normal setup this problem can be settled by changing the convergence criterium to be a relative change in maximum likelihood between steps of less than 0.01% but stata doesn't accept so I had to settle for an ad-hoc criterium of having a norm the gradient of less than 10^{-3} which has the problem of not being scaled relatively to the maximum likelihood.

6.3 Figures
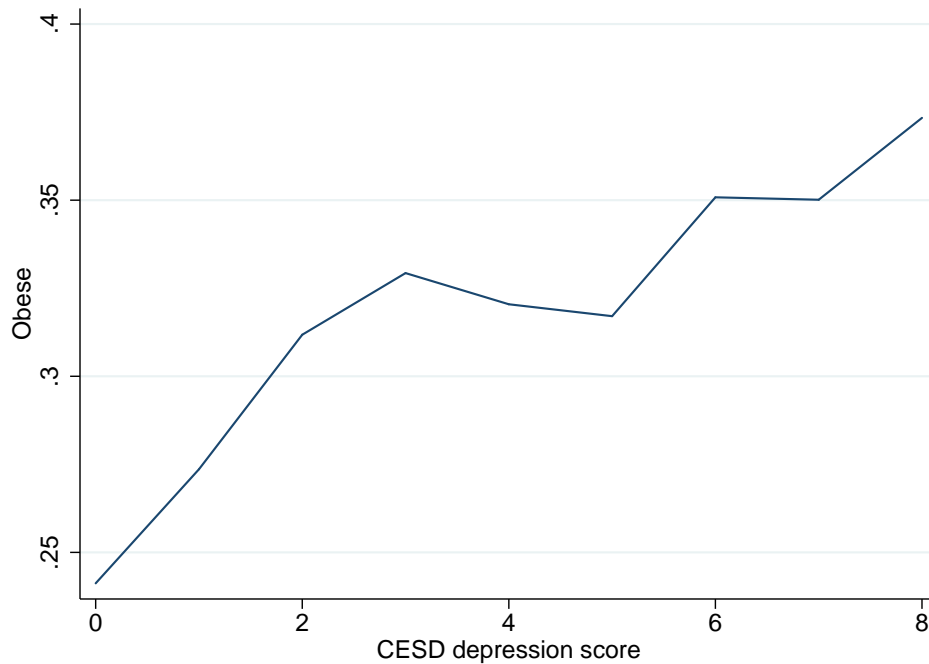
Figure 2: Percentage of obese persons as a function of their depression

Figure 3: Mean depression as a function of the BMI

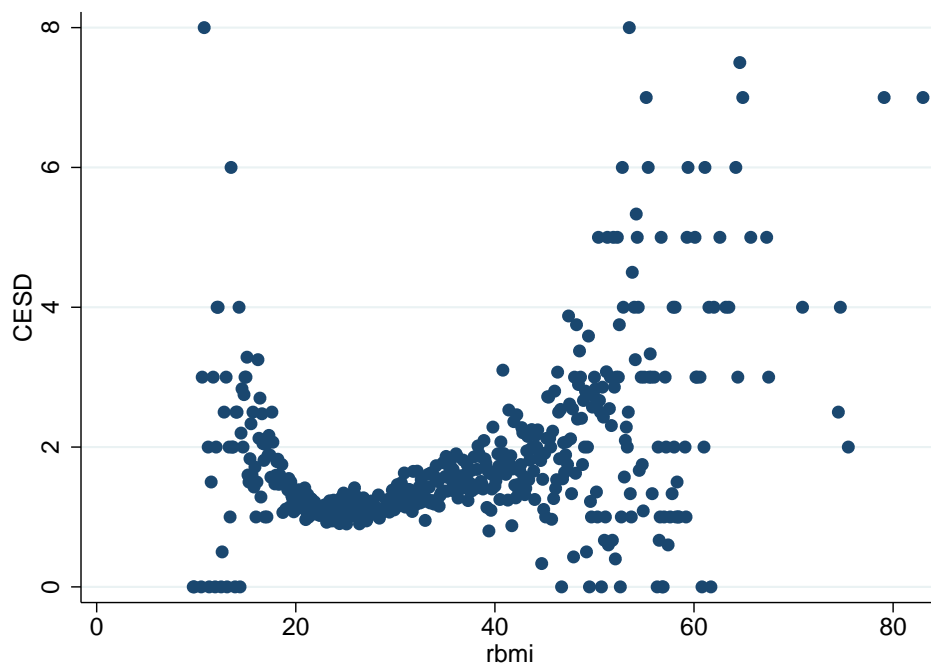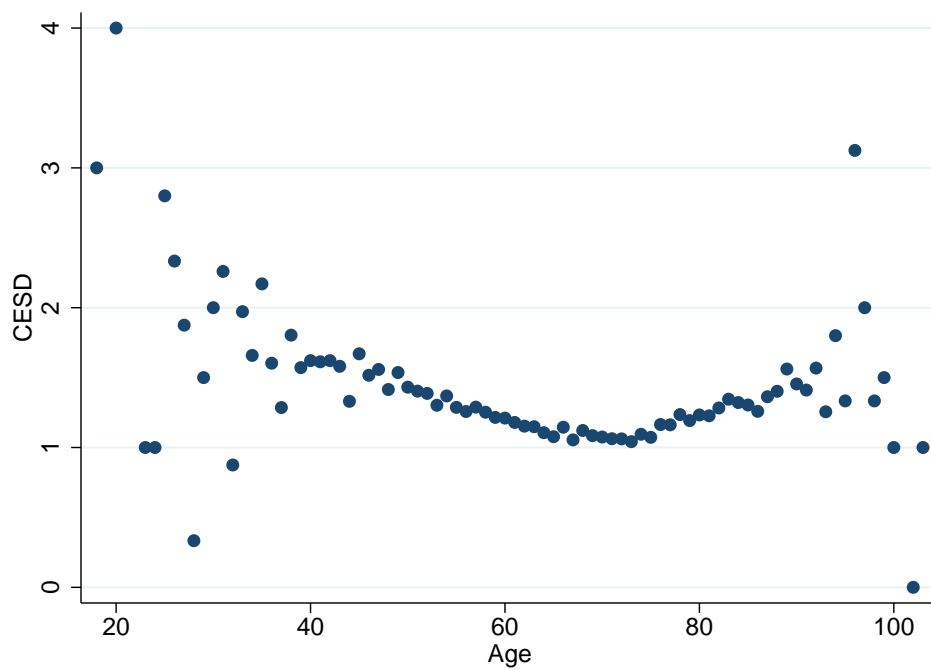Figure 4: Mean depression score as a function of the age
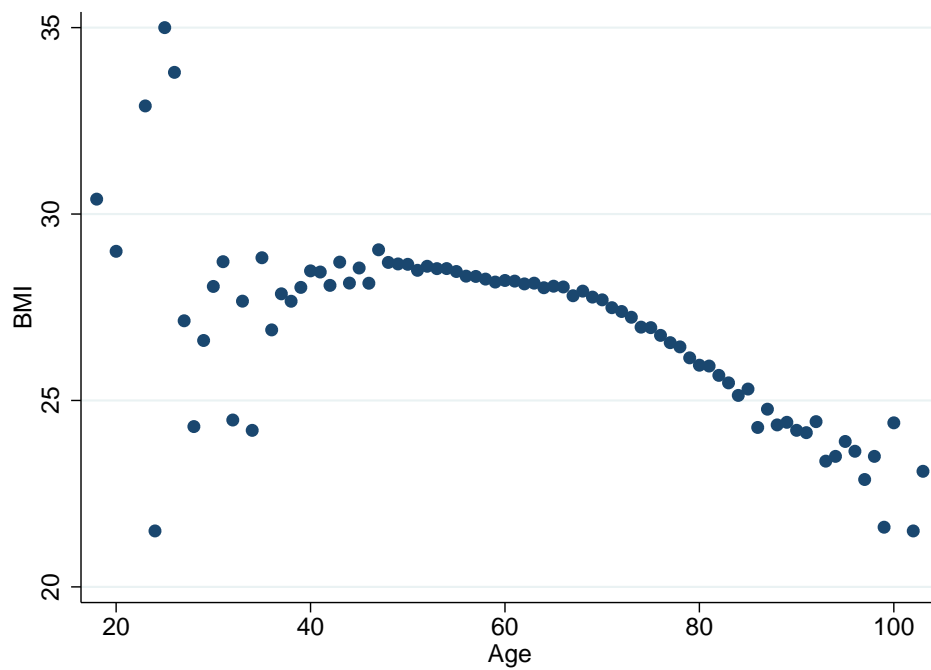
Figure 5: Mean bmi as a function of the age
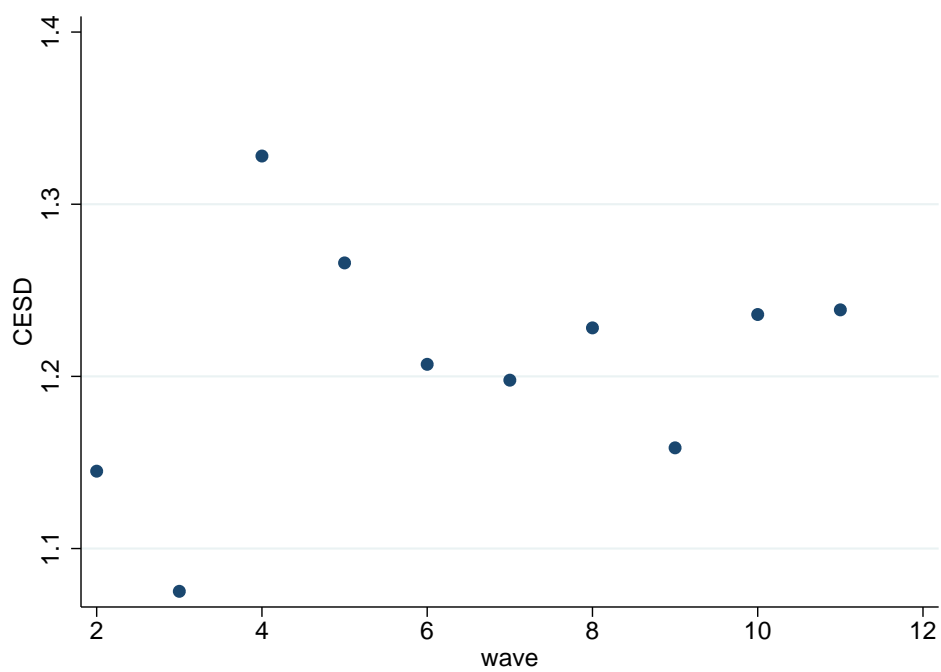
Figure 6: Mean depression score as a function of the wave
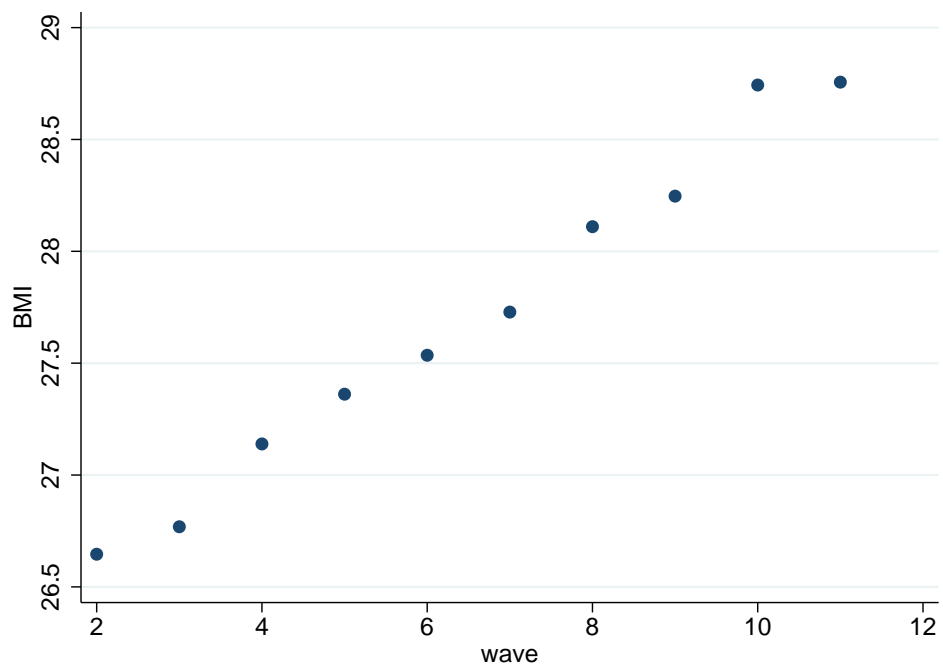
Figure 7: Mean bmi as a function of the wave

6.4 Tables

6.4.1 Descriptive Analysis
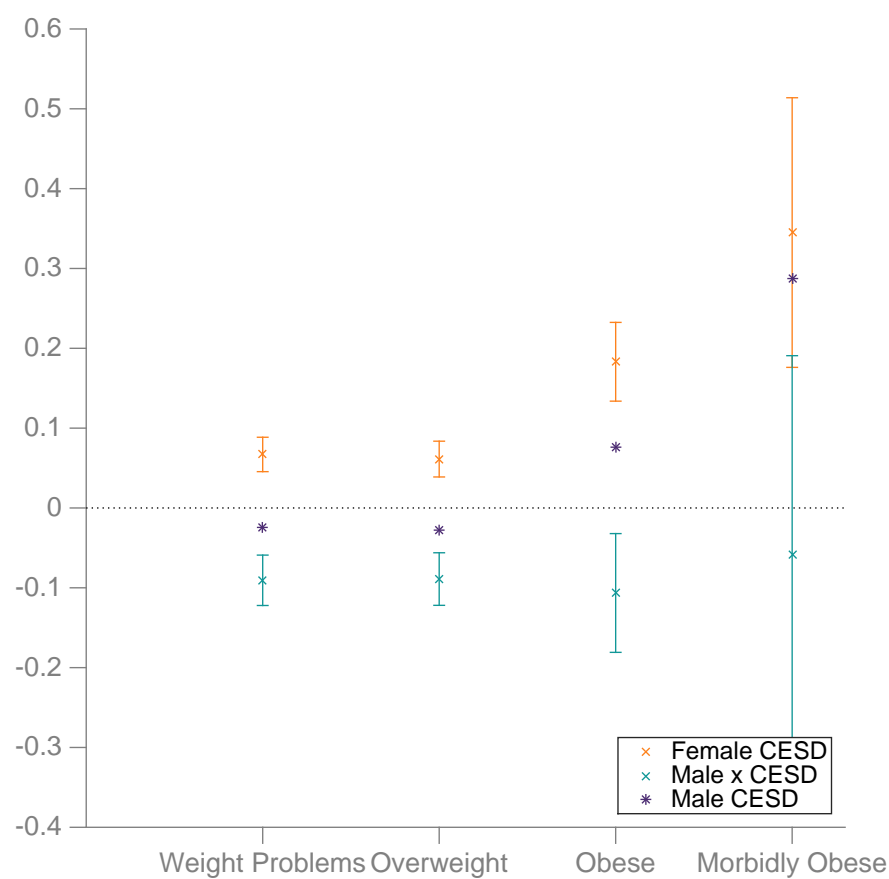
Figure 8: Visual representation of the obtained coefficients for CESD and both sexes, scaled by the proportion of the population in the category

	(1) weightProblems reg	(2) weightProblems ivreg	(3) weightProblems ivprobit marginal
CESD depression score	0.0168***	0.0491***	0.0458***
	(0.00164)	(0.00809)	(0.00751)
Male x CESD	-0.0190***	-0.0665***	-0.0619***
	(0.00244)	(0.0110)	(0.0110)
Male	0.165***	0.221***	0.214***
	(0.00812)	(0.0149)	(0.0137)
has smoked	-0.150***	-0.166***	-0.153***
	(0.0120)	(0.0126)	(0.0109)
Spouse has smoked	0.00324	-0.00109	-0.00137
	(0.00782)	(0.00798)	(0.00788)
Male x has smoked	-0.0288	-0.00317	-0.0150
	(0.0159)	(0.0168)	(0.0154)
Age	-0.00631***	-0.00620***	-0.00605***
	(0.000521)	(0.000523)	(0.000522)
Spouse's Age	-0.000712	-0.000807	-0.000770
	(0.000514)	(0.000516)	(0.000513)
Observations	100600	100600	100600
Education Control	Yes	Yes	Yes
Wave Control	Yes	Yes	Yes

Standard errors clustered by id in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Regression results over the weight problems dependent variable

	(1) overweight reg	(2) overweight ivreg	(3) overweight ivprobit marginal
CESD depression score	0.0152***	0.0445***	0.0412***
	(0.00168)	(0.00825)	(0.00769)
Male x CESD	-0.0190***	-0.0646***	-0.0599***
	(0.00249)	(0.0111)	(0.0113)
Male	0.176***	0.229***	0.222***
	(0.00823)	(0.0151)	(0.0140)
has smoked	-0.176***	-0.190***	-0.175***
	(0.0123)	(0.0128)	(0.0112)
Spouse has smoked	0.00248	-0.00135	-0.00174
	(0.00795)	(0.00810)	(0.00801)
Male x has smoked	-0.0123	0.0121	-0.00267
	(0.0162)	(0.0171)	(0.0156)
Age	-0.00687***	-0.00677***	-0.00659***
	(0.000533)	(0.000535)	(0.000536)
Spouse's Age	-0.000868	-0.000961	-0.000924
	(0.000524)	(0.000525)	(0.000524)
Observations	100600	100600	100600
Education Control	Yes	Yes	Yes
Wave Control	Yes	Yes	Yes

Standard errors clustered by id in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Regression results over the overweight dependent variable

	(1) obese reg	(2) obese ivreg	(3) obese ivprobit marginal
CESD depression score	0.0185***	0.0555***	0.0497***
	(0.00168)	(0.00794)	(0.00683)
Male x CESD	-0.00653*	-0.0317**	-0.0289**
	(0.00259)	(0.0111)	(0.0103)
Male	0.0336***	0.0688***	0.0636***
	(0.00781)	(0.0145)	(0.0137)
has smoked	-0.150***	-0.166***	-0.167***
	(0.00995)	(0.0106)	(0.0113)
Spouse has smoked	0.0124	0.00629	0.00580
	(0.00790)	(0.00799)	(0.00763)
Male x has smoked	0.0123	0.0286*	0.0301
	(0.0134)	(0.0145)	(0.0157)
Age	-0.00720***	-0.00706***	-0.00687***
	(0.000497)	(0.000498)	(0.000479)
Spouse's Age	-0.000264	-0.000291	-0.000445
	(0.000493)	(0.000495)	(0.000473)
Observations	100600	100600	100600
Education Control	Yes	Yes	Yes
Wave Control	Yes	Yes	Yes

Standard errors clustered by id in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Regression results over the obese dependent variable

	(1) morbidlyObese reg	(2) morbidlyObese ivreg	(3) morbidlyObese ivprobit marginal
CESD depression score	0.00619***	0.0165***	0.00997***
	(0.000816)	(0.00330)	(0.00249)
Male x CESD	-0.00152	-0.00890*	-0.00168
	(0.00115)	(0.00403)	(0.00367)
Male	-0.00887***	0.00132	-0.00948
	(0.00255)	(0.00512)	(0.00509)
has smoked	-0.0299***	-0.0344***	-0.0321***
	(0.00369)	(0.00405)	(0.00442)
Spouse has smoked	0.00843**	0.00671*	0.00610*
	(0.00314)	(0.00315)	(0.00273)
Male x has smoked	0.00723	0.0119*	0.00253
	(0.00417)	(0.00470)	(0.00608)
Age	-0.00155***	-0.00151***	-0.00152***
	(0.000190)	(0.000189)	(0.000166)
Spouse's Age	0.0000776	0.0000691	-0.0000860
	(0.000184)	(0.000183)	(0.000164)
Observations	100600	100600	100600
Education Control	Yes	Yes	Yes
Wave Control	Yes	Yes	Yes

Standard errors clustered by id in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Regression results over the morbidly obese dependent variable

	(1)	(2)	(3)
	obese	obese	obese
	reg	ivreg	ivprobit marginal
wave==3	0.0219***	0.0232***	0.0251***
	(0.00342)	(0.00351)	(0.00394)
wave==4	0.0459***	0.0398***	0.0447***
	(0.00401)	(0.00437)	(0.00480)
wave==5	0.0620***	0.0567***	0.0619***
	(0.00438)	(0.00466)	(0.00513)
wave==6	0.0873***	0.0827***	0.0878***
	(0.00483)	(0.00503)	(0.00550)
wave==7	0.0959***	0.0915***	0.0962***
	(0.00490)	(0.00510)	(0.00553)
wave==8	0.126***	0.121***	0.124***
	(0.00525)	(0.00550)	(0.00595)
wave==9	0.146***	0.142***	0.145***
	(0.00560)	(0.00578)	(0.00618)
wave==10	0.161***	0.155***	0.153***
	(0.00539)	(0.00565)	(0.00601)
wave==11	0.173***	0.167***	0.165***
	(0.00561)	(0.00589)	(0.00625)
Observations	100600	100600	100600
Other variables ommitted	Yes	Yes	Yes

Standard errors clustered by id in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Regression explaining obesity only showing the wave control

	(1) obese reg	(2) obese ivreg	(3) obese ivprobit marginal
raeduc==2	0.000891	0.0104	0.00733
	(0.0154)	(0.0156)	(0.0141)
raeduc==3	-0.0196*	-0.000817	-0.00283
	(0.00865)	(0.00973)	(0.00916)
raeduc==4	-0.0452***	-0.0240*	-0.0267**
	(0.00920)	(0.0105)	(0.0100)
raeduc==5	-0.124***	-0.0974***	-0.101***
	(0.00906)	(0.0113)	(0.0116)
Observations	100600	100600	100600
Other variables ommitted	Yes	Yes	Yes

Standard errors clustered by id in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Regression explaining obesity only showing the education control

	(1) obese reg	(2) obese ivreg
CESD depression score	0.0189***	0.0480***
	(0.00132)	(0.00578)
Observations	100600	100600

Standard errors clustered by id in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: Simplest obese regression