The Simultaneous Effects of Obesity, Insurance Choice, and Medical Visit Choice on Healthcare Costs

Colin Baker
National Institute of Health
7021 Wisconsin Ave.
Bethesda, Maryland 20814

Ralph Bradley
Bureau of Labor Statistics
2 Massachusetts Ave. NE
Washington, DC 20212

September 23, 2013

Abstract

While previous studies on obesity's effects on healthcare costs conclude that obesity increases costs, they do not control for the endogeneity of insurance and estimate a tobit model for the corner solution when individuals have no medical expenditure. This study recognizes that there are unobserved heterogeneous factors that guide choices on health insurance, body mass index (BMI) and visiting a provider. Therefore, neither health insurance nor BMI can be treated as exogenous in estimating a cost function and a tobit model must be used to account for corner solutions when the individual does not visit a provider and incurs no medical costs. We find that obesity raises medical costs by \$430.33, and that a 10% reduction in the BMI of each obese person would only lower costs by \$45.28. The obesity elasticity with respect to cost is only .0115%.

1 Introduction

There are ample studies that show the adverse health effects of obesity and its impact on healthcare costs. A publication from the National Institute of Health (1999) cities over 600 medical studies showing that obesity increases the risk of various diseases such as diabetes, stroke, heart, kidney and optic failure. Three examples of studies concluding that obesity increases healthcare costs are Cawley and Meyerhoefer (2012) who conclude "that obesity raises medical costs by \$2,741," Thorpe et. al (2004) and Finkelstein et. al (2003).

Some of the studies such as Finkelstein et. al (2003) and Thorpe et. al. (2004) are correlational and not casual. In these studies, both obesity and body mass index (BMI) are treated as an exogenous variables.[1] However, Cawley and

[1] BMI is derived as $\left(\frac{\text{Weight in Pouns}}{[\text{Height in inches}]^2}\right) \times 703$. The obesity threshhold is a BMI over 30.

Meyerhoefer (2012) recognize that BMI could be an endogenous right side regressor, and use instrumental variable estimation. Their instrument is the BMI of biological children, and thus their study is limited to adults with biological children. In addition, they estimate a two part model for medical expenditures. In the first part, the probability of a non zero medical expenditure is estimated, and in the second part a gamma regression with a log link function is estimated. Insurance status is treated as exogenous.

There are different types of healthcare costs that can be affected by obesity. Like Cawley and Meyerhoefer (2012), we focus on annual per capita costs. Other studies estimate the dollar equivalent cost to the obese from life expectancy loss. Others attempt to investigate the incidence of obesity costs (i.e. who bears the cost - the obese individual or the obese individual's employer). See Bhattacharya and Sood (2011) for a review of this literature.

Despite all the evidence of obesity's adverse health effects, national obesity rates continue to rise even though $60 billion is spent on the private weight loss industry and there are numerous public program interventions coming from National Institute of Health, the US Agriculture Department, and even the White House. This is certainly different from the success at reducing US smoking rates. It is difficult to find reasons that obesity rates continue to climb even though it increases the risk of many diseases. We try to do this with a simple micro model of BMI choice. It has latent variables observable only to the individual that influence preferences when BMI, insurance, and medical visit choices are set. The model predicts that the individual will take into account the BMI choice when making the insurance choice, and conversely, when making the BMI choice, will consider the insurance choice. The endogeneity coming from the latent variables and simultaneously determined choices creates inconsistent estimates unless this endogeneity is properly treated.

Several studies explain the obesity problem through the use of behavioral economics. Ruhm (2012) models weight choice as an interaction between a deliberative (rational) system and an affective system where the weighting of the two systems is a function of an exogenous endowment of "self control". Cutler et. al (2003) suggest that obesity can occur from nonrational discounting of the future benefits of dieting. These behavioral economic models are appealing because they are consistent with a neuroscience based explanation. The difficulty with such models is that there are so many latent variables such as self control and the irrational discount rate that they are hard to verify empirically. We argue in this study that disutility occurs when reducing BMI and the marginal disutility per unit of BMI reduction is randomly distributed across the population. This disutility could easily be a function of an individual's neuro transmitter system, metabolism, access to healthy food, and income/leisure resources to access gyms and weight clubs. While Cawley and Meyerhoefer (2102) emphasize genetics, we argue that genetics is at best only a part of the cause of obesity. BMI is still the result of choices. We use a simple micro model to show that an unhealthy BMI could be a rational maximizing choice where the individual trades off the increased disutility of weight reduction with the increased utility coming from better health. Such a model is still consistent with the behavioral economic approach, and it provides a better guide for econometric specification because it shows where and how the endogeneity occurs.[2]

In this study, instead of always using instruments to correct for endogeneity of BMI, we use control variables as outlined in Newey, Powell and Vella (1999). Unlike Cawley and Meyerhoefer (2102) we do not use a two part model, but instead estimate a multiple selection tobit type model that allows for the possibility that when consumers set their BMI and insurance status, and decide whether to visit a provider, latent variables are common to all these choices. If this is true, then the two part model with exogenous insurance do not provide consistent estimates. Our methods are based on a two period (*ex ante* and *ex post*) micro economic model adapted from Dragone and Savorelli (2012). Their model recognizes that getting one's BMI (through consuming calories below the level of satiation) nearer to an ideal level invokes disutility, and when setting BMI, the consumer must trade off the marginal utility of additional health with the marginal disutility of feeling increasingly unsatiated.[3] In our model, both insurance status and BMI are simultaneously set *ex ante*. After a draw of a random health status variable in the *ex post* period, the consumer chooses whether or not to visit a service provider. If the consumer visits a provider, then based on the consumer's health status, the provider selects a treatment intensity.

This study uses the Medical Expenditure Panel Survey (MEPS) as the data source. One challenge of using MEPS is that if individuals visit a provider such as an emergency room and the provider receives no payment, then the expenditure is recorded as a zero even though the treatment did have an actual cost. Unlike previous studies, this study adjusts for non payers. Another challenge that has not been addressed in the past is that not all MEPS respondents respond to the height and weight questions. If their BMI is a consideration in their non response, there is a problem of nonresponse.

We use eight years of MEPS data from 2002 to 2010. This is a very interesting period to study obesity. During this period MEPS shows that the national obesity rate continues to climb despite the rise in food prices in 2008 and despite the fact that food processing technology does not change as it did during the period of the Cutler et. al. (2003) study. Additionally, during this period, the adverse health effects of obesity are well known. Not only is obesity rising during this period, but the diseases arising from obesity such as diabetes, hypertension, and hyperlipidemia are also rising.

We start this study with a simple micro model that shows that unhealthy BMIs can be a result of an optimizing decision. This model predicts that there can be *ex ante* moral hazard from having health insurance when making BMI choices, and there can be adverse selection where those with greater propensity to have higher BMIs will more likely purchase health insurance.[4] In the simple

[2] Our micro model is consistent with the explanation given by the Center of Disease Control. Their website says, "Body weight is the result of genes, metabolism, behavior, environment, culture, and socioeconomic status." See http://www.cdc.gov/obesity/adult/causes/index.html.

[3] Their focus is on aneroxia nervosa, but their micro model can be easily adapted to obesity.

[4] If wages adjust for the expected *ex post* costs of obesity for all employer plans and if

micro model, the individual has unobserved characteristics that influence the BMI, the insurance decision, the decision to visit a medical provider, and the level of medical expenditures. When these conditions exist, the two part model will not generate consistent estimates.

Since we wish to test the *ex ante* moral hazard and adverse selection predictions of the model, we can only do this when individuals have a health insurance status choice. We therefore limit our analysis to adults who are not eligible for any public insurance program and are free to choose whether or not to be insured.

When we run simulations to estimate the impact of an exogenous reduction in BMI on costs, we use our model first to estimate the effects of BMI change on propensity to insure and propensity to visit a provider. Our final estimate on cost is the sum of the cost impacts due to changes in propensity to insure, medical visitation propensity and the direct effects on costs. The Cawley and Meyerhoefer (2012) estimtes only incorporate the visitation propensity effect and the direct effects on costs.

Unlike previous studies, we account for the endogeneity of BMI by explicitly modelling and estimating BMI choice.[5] Other previous studies have not been concerned with the individual's trade off between the health benefits of a lower BMI with the increased disutility of making the effort to reduce BMI.[6] Since this disutility of effort or BMI outcome cannot be written into an employer sponsored insurance contract (non-contractible) nor can employer sponsored plans risk adjust premiums for the marginal acturial cost of a marginal increase in BMI, we cannot get a "first best" allocation of this disutility. (This is the reason for the *ex ante* moral hazard).[7,8] If genetics is the key factor behind BMI determination and individuals do not choose their own BMI then there is no *ex ante* moral hazard.

We get many interesting empirical results. First, we find evidence that the non ignorable response for the MEPS weight question is most likely for individuals with BMIs between 27 and 30. They are not fully obese, but there is

individual plan premiums adjust for expected *ex post* costs, then there is no *ex ante* moral hazard or adverse selection.

[5]There are other studies in other areas of obesity that also do not account for the endogeneity of obesity. Bhattachayra and Bundorf (2009) estimate the incidence of obesity by running an OLS equation with wage as the dependent and obesity dummies as an exogenous regressor. They get unexpected results such as a positive parameter estimate for the employer sponsor insurance dummy. Many of their obesity parameter estimates are negative but not statistically significant.

[6]We focus on the disutility of BMI reduction because possibly intervention programs misestimate this disutility and make weight reduction sound easier than it is. When participants find that BMI reduction is not as easy as they were lead to believe, they might get discouraged and drop out.

[7]Bhattacharya and Sood (2006) focus entirely on this source of *ex ante* moral hazard, but they do use these words. Instead, they use the words "obesity externality."

[8]Most health plans are employer sponsored and the Health Insurance Portability and Accountability Act requires a pooled premium. Even for individual plans, a marginal BMI addition to premiums can be problematic. Bhattacharya and Bundorf (2009) find that wages for employees with employer sponsored plans do adjust for BMI effects. In this case, there is no *ex ante* moral hazard or adverse selection.

a possibility that it is difficult to assess their weight by appearance. The obese are more likely to report their weight. Second, we find that food prices have no statistically significant effect on BMI choice. Third, we find that an increase in BMI will increase the propensity to purchase health insurance (adverse selection) and the presence of insurance has a positive effect on BMI choice (*ex ante* moral hazard). This confirms the predictions of our micro model. Fourth, there is evidence that there are common latent variables that the researcher cannot observe when individuals make the medical utilization and insurance choices. Thus, correct modelling requires either the use of instrumental or control variables. The fifth finding is not directly related to the effect of obesity on costs and is counter intuitive. Those who have a high propensity not to visit providers on average create more cost because when they are induced to see a provider, their illness has become far more severe and this severity could have been prevented had they seen a provider earlier. Since a higher BMI increases this visit propensity, obesity's effect on this propensity generates a small cost savings. We argue that Cawley and Meyerhoefer focus on the cost of obesity is the wrong focus. However, we find that obesity only increases costs by $430.33 compared to their $2,741. If each obese individual reduces BMI by 10%, on average there will only be $45.28 reduction in medical costs. The cost elasticity of obesity is only .0115%.

Section 2 establishes the micro foundations for the econometric model in this study. Section 3 describes the data and estimation methods, and Section 4 describes the results.

2 A Simple Micro Model

Several micro economic studies employ behavioral economics to explain the presence of obesity. Such studies are Runm (2012) and Cutler et. al. (2003). In this study, we argue that obesity can be the result of a rational utility maximizing process. Our micro model is borrowed from Dragone and Savorelli (2012). While their concern is with anorexia nervousa, it is still useful here because it accounts for the disutility of consuming calories below (or above) a level of satiation. Since body weight is a monotonic function of calories consumed, choosing a calorie consumption is equivalent to choosing a BMI. Therefore, unlike Dragone and Savorelli, we focus solely on the BMI choice.

There are two periods - *ex ante* and *ex post*. In the *ex ante* period, the consumer makes expectations on her health status and medical spending in the *ex post* period. Based on these expectations the consumer decides her insurance status (denoted as I_i where i indexes consumers), and her BMI (denoted as B_i). If the individual decides to buy insurance then $I_i = 1$, otherwise it is 0. Cost sharing respectively under insurance and no insurance is $c_{I,i}$ and $c_{N,i}$ ($c_{I,i} < c_{N,i}$).[9] The ideal BMI (denoted as B_I) does not vary. However, there is a "natural" BMI (denoted as $B_{N,i}$) which occurs when the individual

[9] $c_{N,i}$ can be less than one. Often an uninsured individual can visit a provider and pay nothing for the service. This is particular true of emergency room visits.

eats to satiation and pursues no other activity to manage weight. $B_{N,i}$ varies by individual. The lower the individual's B_i goes below the satiated BMI, $B_{N,i}$, there is an increasing marginal disutility of nonstation. The econometrician cannot observe $B_{N,i}$. When forming expectations, there are characteristics observable by the econometrician (denoted as X_i) and other unobservable characteristics (denoted as ξ_i) that help predict the *ex post* health status (denoted as S_i).[10] When the *ex post* period begins, the consumer draws an unpredictable shock, ε_i and the log of the health status variable is determined by[11]

$$\ln(S_i) = \beta_0 + X_i\beta_1 + (|B_i - B_I|)\beta_2 + \xi_i + \varepsilon_i. \tag{1}$$

S_i measures the severity of the *ex post* illness. A higher S_i indicates a higher illness severity. After the draw of ε_i the individual decides whether or not to visit a service provider. If the individual does visit the provider, the total cost (C_i) is determined as:

$$C_i = Ac_i^{\alpha}S_i,\ c_i \in [c_{I,i}, c_{N,i}], -1 < \alpha < 0. \tag{2}$$

In other words, after making the discrete choice of visiting the provider, total medical cost is set to (2). The individual's out of pocket cost is c_iC_i. The parameter α accounts for any *ex post* moral hazard. The effect on utility from S_i is

$$\begin{aligned} U(S_i) &= -BS_i^{\Gamma},\ \text{no provider visit} \\ &= -BS_i^{\Gamma}C_i^{\theta},\ \text{with a visit}, \Gamma > 2, -1 < \theta < 0. \end{aligned} \tag{3}$$

Medical spending helps lessen the disutility of illness but not fully. To ensure this, the parameter A in (2) is less than B in (3).

The individual visits the provider if the income loss, the nonmonetary cost (t_i),and the disutility of illness after treatment is greater than the disutility of getting no treatment or[12]

$$\begin{aligned} -t_i - c_iC_i - BS_i^{\Gamma}C_i^{\theta} &> -BS_i^{\Gamma},\ \text{or} \\ -t_i - c_iAc_i^{\alpha}S_i - BS_i^{\Gamma}(Ac_i^{\alpha}S_i)^{\theta} &> -BS_i^{\Gamma} \\ -t_i - c_iAc_i^{\alpha}S_i - BS_i^{\Gamma+\theta}(Ac_i^{\alpha})^{\theta} + BS_i^{\Gamma} &> 0 \\ H(S_i, t_i, c_i) &> 0. \end{aligned} \tag{4}$$

The second line is derived by substituting for C_i using (2). At $S_i = 0$, there is no visit since the inequality is not satisfied. However, given the restrictions $\partial H(S_i)/\partial S_i > 0$; as S_i increases there will be a threshold $\overline{S}_i(c_i, t_i)$ where the consumer will be indifferent between visiting and not visiting the provider.

[10] ξ_i is unobservable to the econometrician but observable to individual i.

[11] ε_i is completely unpredictable by the individual and thus, is independent of X_i and ξ_i. Since S_i is a function of unobserved variables, it too is unobservable.

[12] t_i incorporates time costs, anxiety costs and all other non observed nonmonetary costs of seeing a provider.

If $S_i > \overline{S}_i(c_i, t_i)$, the consumer visits the provider. Obviously $\overline{S}_i(c_{I,i}, t_i) < \overline{S}_i(c_{N,i}, t_i)$.

In the *ex ante* period the consumer chooses her B_i and I_i by forming expectations about S_i and the choice to visit the provider in the *ex post* period. She will simultaneously select her BMI and insurance status to maximize the expected utility in the *ex post* period. In terms of the variables known by the consumer, the expected utility conditional on all variables observable to the consumer can be characterized by

$$U = U(\pi_i, c_i, X_i, B_i - B_I, B_i - B_{N,i}, \xi_i). \tag{5}$$

π_i is the insurance premium and is zero if the individual chooses not to buy insurance. c_i is the cost sharing variable and can take on the values $c_{I,i}$, or $c_{N,i}$ depending on the insurance choice. The fourth argument measures the impact on expected utility from deviating from the ideal BMI, B_I. As (1) shows, a greater deviation from the ideal BMI increases expected illness severity.[13] The fifth argument measures the disutility of deviating away from the individual's natural BMI, $B_{N,i}$. It accounts for the increasing disutility of non satiation (and discomfort of physical activity) as the consumer moves further away from her natural BMI. Let U_j and U_{jk} be respectively the first derivative with respect to the j^{th} argument and the second derivative with the j^{th} and k^{th} argument. Suppose that $B_{N,i} > B_I$, and B_i is any value between $B_{N,i}$ and B_I, and the following holds[14]

$$\begin{aligned} B_i &> B_I \Longrightarrow U_4 < 0, U_{44} < 0 \\ B_i &= B_I \Longrightarrow U_4 = 0, U_{44} < 0 \\ B_i &< B_{Ni} \Longrightarrow U_5 > 0, U_{55} < 0 \\ B_i &= B_{Ni} \Longrightarrow U_5 = 0, U_{55} < 0 \\ U_1 &< 0 \\ U_2 &< 0,\ U_{24} < 0. \end{aligned} \tag{6}$$

The fourth argument of (5) is maximized when $B_i = B_I$ for any fixed values of the other arguments The fifth argument it is maximized when $B_i = B_{Ni}$ for any fixed values of the other arguments. Since $B_{N,i} > B_I$, if the consumer reduces B_i there is marginal increase in utility from the fourth argument, but a marginal decrease in the fifth argument. For a fixed insurance status, when the consumer selects B_i and $B_{N,i} > B_I$, there is a trade off between all the benefits coming from improving health and suffering the disutility of deviating from the natural BMI.

[13] An increase of B_i away from B_I increases S_i. (3) gives the reduction in utility from this additional severity, and (4) gives the income loss from increased medical expenditures. There could be other sources of utility loss such as reduced income from productivity losses and non monetary costs of increased social disapproval. Bhattachary and Sood (2006) give more detail than this study on the results of income loss from productivity loss.

[14] $U_{24} < 0$ occur because as c increases the *ex post* financial impact of a higher level from illness resulting from the increased BMI increases. If wages adjust for the actuarial cost of increased BMI. then $U_{2,4} = 0$.

Given the conditions in (6), it is easy to see that if $B_{N,i} > B_I$, then optimal B_i choice will be in the strict interior of the interval, $[B_I, B_{N,i}]$. To see this, for any π_i, c_i, X_i, ξ_i, the first order conditions for the optimal B_i are $U_4 + U_5 = 0$. If B_i equals either B_I or $B_{N,i}$, the first order conditions fail. If $B_i = B_I$, then individual i can increase expected utility by increasing B_i.[15]

The optimal B_i is also increasing in $B_{N,i}$. A total differentiation of the first order conditions with respect to B_i and $B_{N,i}$ gets

$$\begin{aligned} (U_{44} + U_{55} + 2U_{54})dB_i - U_{55}dB_{N,i} &= 0 \\ \frac{dB_i}{dB_{N,i}} &= \frac{U_{55}}{U_{44} + U_{55} + 2U_{54}} > 0. \end{aligned} \tag{7}$$

This result shows that $B_{N,i}$ can be high enough that obesity is a rational and optimal choice.

If the consumer decides to purchase health insurance then the first order conditions for the optimal choice of B_i is $U_4(\pi_i, c_{I,i}, X_i, B_i - B_I, B_i - B_{N,i}, \xi_i) + U_5(\pi_i, c_{I,i}, X_i, B_i - B_I, B_i - B_{N,i}, \xi_i) = 0$, and if the consumer decides not to buy insurance, the first order conditions are $U_4(0, c_{N,i}, X_i, B_i - B_I, B_i - B_{N,i}, \xi_i) + U_5(0, c_{N,i}, X_i, B_i - B_I, B_i - B_{N,i}, \xi_i) = 0$. Letting B_i^{I*} and B_i^{N*} be respectively the optimal choices for BMI for being insured and uninsured, the consumer choose to be insured if

$$\begin{aligned} &U(\pi_i, c_{I,i}, X_i, B_i^{I*} - B_I, B_i^{I*} - B_{N,i}, \xi_i) > \\ &U(0, c_{N,i}, X_i, B_i^{N*} - B_I, B_i^{N*} - B_{N,i}, \xi_i). \end{aligned} \tag{8}$$

Given the conditions in (6), we show in section A.1 of the appendix that $B_i^{I*} > B_i^{N*}$. Thus, insurance can generate *ex ante* moral hazard when it comes to BMI choices.[16]

In section A.2 of the appendix, we show that there is also adverse selection or equivalently that an increase in $B_{N,i}$ increases the propensity to purchase health insurance.

This simple model predicts both *ex ante* moral hazard and adverse selection. The empirical section of this study will test the predictions of this simple micro model. The intuition here is that $B_{N,i}$ is private, asymmetric information that only the individual knows. The premium, π_i, cannot be risk adjusted for this private information. Since the optimal B_i^* choice monotonically increases with $B_{N,i}$, we can use B_i^* as an endogenous proxy when econometrically testing

[15] The second order condition is $U_{44} + U_{55} + 2U_{54} < 0$.

[16] There are two types of moral hazard, *ex post* and *ex ante*. *Ex post* moral hazard occurs from the *ex post* over consumption of medical services because the consumer does not pay the full marginal costs. *Ex ante* moral hazard occurs because efforts to prevent diseases are non contractible in an insurance policy or premiums can't adjust for BMI choics and consumers are not compensated for the effects that their efforts at prevention have on expected benefits. Additionally, they get a lower return on their preventive efforts because when they they are only paying a fraction of the full costs of getting ill. In this study the effort is the disutility of non satiation when setting the BMI below the natural BMI. See Bradley (2005) or Bhattacharya and Sood (2006) on a fuller depiction of *ex ante* moral hazard and Pauly (1968) on *ex post* moral hazard.

for adverse selection. The *ex ante* moral hazard occurs because the insured individual bears a smaller financial impact for her BMI decisions, and the BMI choice cannot be written into a health insurance contract.

In this framework, insurance choice, BMI choice, provider visits, and medical costs are influenced by variables $B_{N,i}$ and ξ_i that are not observable to the econometrician. Simply modelling medical cost (or C_i in (2)) by using insurance and B_i as an exogenous regressors, and not accounting for the provider visit decision will lead to endogeneity bias. Obviously, we cannot estimate (2) directly because we cannot observe S_i. When we substitute (1) to (2) and take logs the estimating cost equation is

$$\ln C_i = a + \alpha \ln(c_i) + X_i\beta_1 + (|B_i - B_I|)\beta_2 + \xi_i + \varepsilon_i. \tag{9}$$

The coefficient of interest is β_2. However, we cannot observe ξ_i, Yet, it influences both insurance and BMI choice. Suppose both $C_i > 0$ and $I_i = 1$, then both conditions (4) and (8) hold where $H(S_i(X_i, \xi_i), t_i, c_i) > 0$, and $U(\pi_i, c_{I,i}, X_i, B_i^{I*} - B_I, B_i^{I*} - B_{N,i}, \xi_i) > U(0, c_{N,i}, X_i, B_i^{N*} - B_I, B_i - B_{N,i}, \xi_i)$. The the right side regressors of (9) are correlated with

$$\begin{gathered}\lambda(\pi_i, c_i, t_{i,} X_i) = \\ E(\xi_i | \{H(S_i(X_i, \xi_i), t_i, c_i) > 0\} \\ \cap \{U(\pi_i, c_{I,i}, X_i, B_i^{I*} - B_I, B_i^{I*} - B_{N,i}, \xi_i) > \\ U(0, c_{N,i}, X_i, B_i^{N*} - B_I, B_i - B_{N,i}, \xi_i)\}).\end{gathered} \tag{10}$$

We rewrite (9) as:

$$\begin{aligned}\ln C_i &= a + \alpha \ln(c_i) + X_i\beta_1 \\ &\quad + (|B_i - B_I|)\beta_2 + \lambda(\pi_i, c_i, t_{i,} X_i) \\ &\quad + \nu_i + \varepsilon_i, \\ \xi_i &= \lambda(\pi_i, c_i, t_{i,} X_i) + \nu_i.\end{aligned} \tag{11}$$

Here (11) is a tobit model with two selection effects, the insurance decision and the provider insurance effects. Models such as these are rarely covered in the econometrics literature. Maddala (1983 pp 278-283) briefly covers models with multiple selectivity and without proof provides the estimating procedure for extending the heckit model for two selection effects.

The micro foundations in this section lead me to a different estimation strategy than Cawley, J. and Meyerhoefer (2012), who emphasize evidence that genetical factors are the major determinant of weight. They do not model BMI determination as the result of decisions based on unobserved conditions. However, our micro model predicts that ex post medical costs are a function of the simultaneous ex ante insurance and BMI decisions. BMI decisions will affect costs both directly and through health insurance decisions. This is a feature that the Cawley, J. and Meyerhoefer (2012) model misses. In our model, the natural BMI, $B_{N,i}$ is a condition that could easily be influenced by genetical factors, but in the end, individual i's BMI, B_i, is the result of a decision making process. To correct for the endogeneity of BMI, we use a control variable approach where we estimate a reduced form equation for B_i. To test the result, $B_i^{I*} > B_i^{N*}$ or that insurance induces the increase in BMI we estimate a structural form for B_i where private insurance is endogenous.

This study's biggest departure from Cawley, J. and Meyerhoefer (2012) is that they use a two part model where in the first part the provider visit decision is estimated with a logit model and in the second part, the cost equation conditional on non zero medical expenditures is estimated independently as a gamma regression with a log link. They do not mention how they treat insurance choice and they do not even report insurance status as a summary statistic. Their methods will only provide consistent estimates as long as the multiple selection effects, $\lambda(\pi_i, c_i, t_{i,} X_i)$ in (11) are zero everywhere. We find that the multiple selection effects are statistically significant.[17]

3 Data and Estimating Equations

3.1 Data

The Medical Expenditure Panel Survey (MEPS) is stratified random sample of households in the US where each household remains in the sample for two years. Each year new households are sampled and for a given year, a household was either in the sample in the previous year or it was not. The survey collects for each household individual her medical expenditures, her diagnosed diseases, her perceived health status, her insurance status, her employment and her demographic variables. While each household is interviewed five times, medical expenditures are only reported annually. This survey also surveys the medical providers and pharmacies used by the households in order to obtain more accurate expenditure data.

The data has several files. The household file has each individual as a unique observation and lists the total annual medical expenditure along with the economic, demographic and BMI information. The conditions file has a diagnosed condition as the unique observation and a new record is created with each newly reported treated disease. The event files has a separate record for each office, outpatient, emergency room and hospital visit. There is also a separate record for each pharmaceutical refill.[18]

Since 2002 MEPS has collected individual BMIs. Thus, the sample in this study starts in 2002 and ends in the last year available, 2010.

Figure 1 shows how the obesity rate has climbed from 2002 to 2002. Figure 2 compares the kernel densities for BMI for 2002 and 2010. The 2010 distribution is "flatter" and mass migrated from the 21 to 26 range in 2002 to the 30 to 45

[17] There is a debate between the relative merits of the two part model and the heckit model. Dow and Norton (2003) argue that the heckit model are often misused, and t-tests for the null hypothesis, $\lambda(\pi_i, c_i, t_{i,} X_i) = 0$ perform poorly.

[18] Since MEPS is stratified sample, consistent variance estimation requires accounting for the clustering of the primary sampling units and strata.

range in 2010. Figure 3 compares the kernel densities for nominal per person medical expenditures. Both the 2002 and 2010 distributions are skewed to the left. Again the 2010 distribution is flatter and there are larger outliers. Figure 4 compares individual medical expenditures in 2002 dollars.[19] While there is a slight increase in the average, the densities have not changed greatly.

Table 1 lists selected summary statistics for the beginning year and ending year of this study.[20] Of note the national obesity rate has climbed from 17.5% in 2002 to 21.5% in 2010. The fraction of individuals with no medical visits and cost ($C_i = 0$) rose from 14.8% in 2002 to 15.4% in 2010. MEPS attempts to record actual household expenditures. If an individual visits, say an emergency room, and this visit is not reimbursed then the event file will record a zero expenditure for this visit. In 2002 14.7% of all individuals had at least one fully unreimbursed visit and this rose to 15.3% in 2010. This can present challenges when attempting to measure the effect of obesity on costs. Even though a visit goes unreimbursed, this does not mean that the cost of the visit is zero. This table also shows that in 2010 a smaller fraction of the population had access to a primary care physician and were covered by private health insurance. Nominal medical spending per capita increases from $2,813 to $4,094 while medical spending in 2002 dollars rises only to $3010 in 2010.

We impute a cost for unreimbursed visits by a using an average with a shock for reimbursed expenditures. While there are controversies behind this approach we do it to get an alternative per capita cost. Table 1 shows that the imputation in 2002 adds $600 to the per capita costs.

3.2 Estimating Equations

3.2.1 *Ex Ante* Period

To simplify exposition, we change notation slightly in this subsection. Now X represents all the observable right side covariates, and ξ represents all unobservables, including $B_{N,i}$ and t_i. Different models will have different right side covariates and there is an additional subscript to distinguish the different covariates among the different models.

We model the *ex ante* period first where BMI and insurance choices are made. We want to verify that there is both ex ante moral hazard and adverse selection as the simple micro model predicts.

We start with the reduced form BMI (B_i) equation. We first need to control for the effects of responding to the weight question as not all MEPS respondents responded. We estimate a probit model for responding to the weight question in MEPS. Let $X_{i,R}$ be the observable characteristics that govern the response

[19] To get real medical expenditures we deflate by the Medical CPI so that all medical expenditures can be expressed in 2002 dollars.

[20] The standard errors of the mean are in parenthesis. Since MEPS is stratified random sample, variance estimation needs to account for the stratification. In this study we use the Taylor Series (linearization) method.

to the weight question in MEPS. The individual responds if

$$X_{i,R}\beta_R + u_{i,R} > 0 \tag{12}$$

where $u_{i,R}\tilde{}N(0,1)$ and contains the effects of ξ_i. (From here on, all residuals, u, contain the effects of ξ_i; so for (12) $u_{i,R} = \gamma_R\xi_i + \nu_{R,i}$ where $\nu_{R,i}$ is an unobservable residual that effects the response decision but not the *ex post* variable S_i.) Letting $\widehat{\beta}_R$ be the parameter estimate. We can next estimate the reduced form equation for BMI (B_i) as

$$B_i = X_{i,B}\beta_B + \lambda(X_{i,R}\widehat{\beta}_R) + u_{i,B} \tag{13}$$

where $u_{i,B}$ is a mean zero residual and $X_{i,B}$ are exogenous covariates. $\lambda(X_{i,R}\widehat{\beta}_R)$ is the inverse mills ratio using the parameter estimate from (12). The estimated residual, $\widehat{u}_{i,B}$, is the control variable that corrects for the endogeneity of B_i in the other models.

We estimate a structural health insurance choice model using the control variable to correct for the endogeneity of BMI choice. Let $X_{i,I}$ represent both the observable endogenous and exogenous covariate influencing insurance choice. Then

$$\begin{aligned} I_i &= 1 \Longrightarrow X_{i,I}\beta_I + u_{i,I} > 0 \\ I_i &= 0 \Longrightarrow X_{i,I}\beta_I + u_{i,I} \leq 0. \end{aligned}$$

In this model the coefficient for BMI is the coefficient of interest. If it is positive and significant, this gives evidence that there is adverse selection.

Next, we add private insurance status to $X_{i,B}$ in (12) and account for its endogeneity. When we re estimate this BMI model, interest is on the private insurance coefficient. The coefficient of interest is the private insurance effect. If it is positive and significant, then there is evidence of *ex ante* moral hazard.

3.2.2 *Ex Post* Period

In the ex post period, the individual decides to whether or not to visit a provider, and if there is a visit then medical expenditures are set. As discussed in the section in the micro model, the BMI, insurance, and provider visit decisions are a function of unobserved individual characteristics, ξ_i. The residuals, $u_{i,B}$ and $u_{i,I}$, from the ex ante models are functions of ξ_i as they were in the *ex ante* subsection.

The decision to visit a medical provider and the resulting medical expenditure from a visit are also functions of ξ_i. Therefore, BMI is an endogenous right side regressor where a control variable is used to correct for its endogeneity. The ex ante choice of insurance status, and the ex post decision to visit a provider generate a multi selection effect. The individual decision to visit a provider is specified as

$$\begin{aligned} C_i &> 0 \Longrightarrow X_i^C\beta_C + u_{i,C} > 0 \qquad (14) \\ C_i &= 0 \Longrightarrow X_i^C\beta_C + u_{i,C} \leq 0. \end{aligned}$$

Finally, if $C_i > 0$, then medical expenditures estimated as a gamma regression with mean μ_i and a log link function

$$\ln \mu_i = X_i^{C>0} \beta_{C>0}$$
$$+\text{multiple selection effects.}$$

Notice that $X_i^C \beta_C$ in the visit choice equation (14) is not the same as $X_i^{C>0} \beta_{C>0}$ in the medical cost equation since it is the individual that is solely involved in the visit decision, but the physician is involved in the setting of medical expenditures. In appendix A.2, we detail how we first estimate the multivariate probit for the joint event of being insured and visiting a provider or

$$\Pr(\{X_i^C \beta_C + u_{i,C} > 0\} \cap \{X_i^I \beta_I + u_{i,I} > 0\}),$$

and then use this estimation to compute the multiple selection effects.

4 Results

4.1 Models for the *ex ante* period

Table 2 lists the parameter estimates of the probit model in (12) for responding to the MEPS weight questionnaire. Males are more likely to respond than females. Response improves with education. Most of the year dummies do not produce significant results. Unemployed individuals are less likely to respond. As one ages, one is less likely to respond.

Table 3 lists the results for the reduced form BMI equation in (13). We used the PPI Price Index for corn syrup divided by the all items CPI as a proxy for the relative price for the food. Corn syrup is an intermediate product for foods considered the major culprit behind obesity. The parameter estimate is negative but not significant. Income is excluded for endogeneity reasons but we include a regressor label "Number of High Occupations." This is the total number of people in the individual's household who are either in a professional, technical, or government occupation. It proxies one's ability to access resources that can help control weight such as gyms, and better food. We exclude the individual's own income because of possible income discrimination against obese individuals. We include spouse's income, and is set to 0 for single individuals. While the coefficient on the Number of High Occupations is significantly negative, the coefficient for the spouse's income is positive and significant although it is small in magnitude. The most interesting result is that if one does a simple heckit the coefficient on the Inverse Mills ratio is negative. We then add the square of the Inverse Mills ratio. The parameter estimate for the squared term is negative while it is positive for the regular Inverse Mills ratio. It seems that the BMI's where the sum of these two terms peak is in the 26 to 28 BMI range. This is the range where it is perhaps most possible to hide one's true weight.

Table 4 shows the parameter estimates for the *ex ante* insurance choice in (14). The coefficient of interests is for the BMI (B_i) and it is significantly positive. This leads to the conclusion that there is adverse selection with BMI.

Individuals with higher BMI are more likely to purchase insurance. The other results are not surprising. Young men have a lower propensity to purchase insurance where as individuals with children who do not benefit from the State Childrens Health Insurance Program (SCHIP) where both spouse work in technical, professional, or government occupations have a much higher propensity to purchase insurance.

Table 5 lists the parameter estimates from a structural BMI equation where private insurance is treated as an endogenous variable. The coefficient of interest is the dummy variable for being privately insured. This provides evidence of *ex ante* moral hazard.

In the *ex ante* period, both insurance status and BMI are determined. If the individual purchases insurance, the financial consequences of illness are less severe, and the policy holder is not compensated by the plan for the savings generated by suffering additional disutility to get the BMI nearer to an ideal level. This is *ex ante* moral hazard.

Likewise, employer sponsored insurance premiums do not seem to be risk adjusted for increases in BMI. As BMI increases so does the risk of severe diseases. This increases the expected utility of holding health insurance. This is adverse selection.

4.2 Models for the *ex post* period

The goal is to estimate a cost equation. Yet, the choice to visit a provider in the *ex post* period and the *ex ante* choice of insurance status are statistically dependent decisions (because of ξ_i) and will influence medical spending if and when the individual decides to visit a provider.

We estimate a multiple selection model where the estimation methods is detail in section A.3 of the appendix. In this method, we first estimate a bivariate normal probit for insurance choice ($I_i = 1$ or 0), and for provider choice ($C_i > 0$ or $C_i = 0$). The results of this model are detailed in Table 6. The income variables have been scaled where they are divided by \$100,000. The estimated parameters have signs that are expected except for the "Poor Perceived Health" coefficient which is negative in the insurance choice equation. Perhaps, most who have poor perceived health find that medical treatments are not effective at mitigating their illness and this gives them less propensity to insure. The coefficient ρ that measures the statistical dependence between the two decisions is positive and significant.

It should be noted that the corner solution tobit effects for ($C_i > 0$) are not as simple as the Standard Tobit Model (Type 1) as depicted in 10.2 of Amemiya (1985). It better conforms to the Type 2 definition as defined in section 10.7 of Amemiya (1985) where the covariates of the selection effect of choosing to visit can be different from the covariates in the medical expenditure equation. The visiting decision is made solely by the individual where as the physician has final authority over the medical expenditure decision.

The parameter estimates for the gamma medical expenditure regression with a log link function are listed in Table 7. We estimated one regression without imputing the zero costs for unreimbursed payment and another with the imputed costs. All the coefficients have the expected sign except for the visit selection effect, $E_{i,2}$ as defined in (21) of section A.3 of the appendix. The result here says that those with a very low propensity to visit a provider do end up generating higher costs when they do see a provider. In our micro model in (4), the individual's underlying illness severity, S_i is not the only variable influencing the decision to visit the provider. There is also a non monetary cost variable, t_i, that is randomly distributed throughout the population. This variable will have a large absolute value if individual i has a phobia against visiting providers. Suppose that individuals i and j have the same observable covariates, X, but if i has a phobia against visiting physicians but j does not, then $t_i > t_j$. This implies that threshold sickness level of i to visit a provider is greater that j's threshold level. Since they have the same observable variables, it must be that when i does visit a provider the expected value of ξ_i is greater that the expected value of ξ_j. This implies that given that both i and j have decided to visit a provider, the expected value of i's expenditure will be greater than j's expected expenditure. For example, suppose both individuals have colon cancer. Individual j goes to the provider when this cancer is in its early stages, and individual i waits until the cancer is extreme and has spread throughout his body.

The other coefficients have their expected signs. The insurance effect is positive and significant as expected. The BMI coefficient is significant and positive.

4.3 Simulations

We run three separate simulations. The first one estimates the cost of obesity on a per person basis. This is the same estimation as the $2,741 estimate made by Cawley and Meyerhoefer (2012). The second one estimates the effect of a 10% BMI reduction for all obese persons. The last estimates the obesity elasticity of cost.

The results of the simulation are listed in Table 8, our counterpart estimate to Cawley and Meyerhoefer (2012) is $430. We break down the components of this effect into the effects coming from insurance change and change in visit propensity as well as the direct effect. Notice that the increased visiting propensity actually reduces costs be $4. Our result represents 14% of real per person expenditures in 2010. My results are 84% lower than Cawley and Meyerhoefer (2012).

Obesity has always been with us and it will not go away. Therefore, we do not believe that the correct question is the cost of obesity. It might be more instructive to determine the impact of an exogenous 10% in BMI for all obese persons. Table 8 reports a $45 reduction if all obese persons reduce their BMI by 10%. There are many reasons that our results might differ from the Cawley and Meyerhoefer (2012) results. Our estimation uses all adults who are not eligible for public insurance while they use only adults with biological children. Our estimation methods are vastly different. We use a control variable method to account for the endogeneity of BMI in a cost estimation, they instrument with the BMI of biological children. We also model and estimate how individuals make their BMI decisions and this influences our parameter estimates, but do not influence Cawley and Meyerhoefer's (2012) estimates. We also account for the endogeneity of insurance.

Finally, we find the percent reduction in costs for every 1% decrease in BMI for obese person. Here the elastisticity is only .0115%.

High BMI does increase costs, but a policy that is successful in reducing BMI will not generate the cost savings that were previously thought.

5 Conclusions

While we do find that obesity does have a positive impact on healthcare costs, its magnitude is more lower than that of Thorpe et. al. (2004) and especially Cawley and Meyerhoefer (2012). It more confirms results from Baker and Duchnovny (2010).[21] They found that "if the distribution of adults by weight between 1987 and 2007 had changed only to reflect demographic changes, then health care spending per adult in 2007 would have been roughly 3% below the actual 2007 amount."

Nonetheless, obesity is a national problem and it continues to increase. While we have found that moral hazard plays a role in setting BMI choices and likewise BMI is a consideration in health insurance choices, we have not been able to answer the questions, "Why is obesity increasing when we know its adverse health effects?" and "Why haven't past private and public interventions worked?" The answers to these questions perhaps require the coordinated research of many disciplines - biology, epidemiology, statistics, and maybe even economics. Yet, our micro model might provide an initial clue. Perhaps, current intervention programs under estimate the marginal disutility that obese individuals face when reducing an additional BMI. People enter these interventions with a false notion of the require effort, and this leads most to fail.

One major problem of modelling and estimating healthcare costs is that the observable covariates such as age, gender, race, etc. explain very little of the variation of health care costs. This gives evidence that the unobserved characteristics that we denote as ξ_i in this study play a larger role in cost determination that the observable characteristics.

We have findings that are unrelated to obesity, but they are important. A higher propensity to visit a provider reduces expected health costs because diseases can be treated at an earlier stage. Important in this decision to visit a provider is the access to a primary provider. The MEPS survey shows that from 2002 to 2010 the percentage of individuals with a primary provider has dropped from 79.7% to 78.0%. This is a disturbing trend and could have negative effects on both future costs and health outcomes.

[21] Baker C. and Duchnovny N., Congressional Budget Office, (2010), "How Does Obesity in Adults Affect Spending on Health Care?" *Economic and Budget Issue Brief*, September, 8 2010.

References

[1] Amemiya, T., (1985), *Advanced Econometrics*, Harvard University Press, Cambridge MA.

[2] Baker C. and Duchnovny N., (2010), "How Does Obesity in Adults Affect Spending on Health Care?" *Economic and Budget Issue Brief*, Congressional Budget Office, September, 8 2010.

[3] Bhattacharha, J. and Bundorf, M.K., (2009), "The Incidence of the Healthcare Cost of Obesity," *Journal of Health Economics*, 28, 649-658.

[4] Bhattacharha, J. and Sood, N. (2011), "Who Pays for Obesity," *Journal of Economic Perspectives*, 25(1), 139-158.

[5] Bhattacharha, J. and Sood, N. (2006), "Health Insurance and the Obesity Externality," *Advances in Health Economics and Health Services Research*, eds. K. Bolin and J. Cawley.

[6] Bradley, R., (2005), "A Semi-Parametric Model For Asymmetric Information In Health Plan Markets," *Economic Inquiry*, 43: 812–822.

[7] Cawley, J. and Meyerhoefer, C., (2012), "The Medical Care Costs of Obesity: An Instrumental Variables Approach," *Journal of Health Economics*, 31(1), 219-230.

[8] Cutler, D.M., Glaeser, E.M., and Shapiro, J.M., (2003), "Why Have Americans Become More Obese?", *Journal of Economic Perspectives*, 17(3), 93-118.

[9] Dow W.H. and Norton E.C.,(2003), "Choosing Between and Interpreting the Heckit and Two-Part Models for Corner Solutions," Health Services and Outcomes Research Methodology, 4, 5-18.

[10] Dragone, D. and Savorelli, L. (2012) "Thinness and Obesity: A Model of Food Consumption, Health Concerns, and Social Pressure," *Journal of Health Economics*, 31(1), 243-256.

[11] Finklestein, E.A., Fiebelkorn, I.C., Wang G., (2003), "National Medical Spending Attributable to Overweight and Obesity: How Much and Who is Paying?" *Health Affairs*, W3-W219.

[12] Maddala G.S., (1983), *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, New York.

[13] Manjunath B.G. and Stephan W., (2012), "Moments Calculations for the Doubly Truncated Multivariate Normal Density," Working Paper, University of Portugal.

[14] National Institute of Health, (1999) "NHLBI Obesity Education Initiative. Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults." Publication Number No. 98-4083.

[15] Newey, W.K, Powell J.K and Vella F., (1999),"Nonparametric Estimation of Triangular Simultaneous Equations Models," Econometrica, 6793), 565-603.

[16] Pauly, M. V., (1968) , "The Economics of Moral Hazard: Comment," *The American Economic Review*, Vol. 58, No. 3, Part 1, 531-537.

[17] Ruhm, C.J. (2012), "Understanding Overeating and Obesity," Journal of Health Economics, 31(6), 781-798.

[18] Thorpe K.E., Florence, C.S., and Joski P., (2004), "Which Medical Conditions Account for the Rise in Health Care Spending?" Health Affairs, W4.437, 437-445.

A Appendix

A.1 Proof $B_i^{I*} > B_i^{N*}$

This is the proof that $B_i^{I*} > B_i^{N*}$. Differentiating $U_4(0, c_{N,i}, X_i, B_i - B_I, B_i - B_{N,i}, \xi_i) + U_5(0, c_{N,i}, X_i, B_i - B_I, B_i - B_{N,i}, \xi_i) = 0$ with respect to both B_i and c_i gets

$$U_{42}dc_i + (U_{44} + U_{55} + 2U_{54})dB_i = 0.$$

The second order condition of the optimization for B_i is $U_{44} + U_{55} + 2U_{54} < 0$. Thus

$$\frac{dB_i}{dc_i} = -\frac{U_{42}}{U_{44} + U_{55} + 2U_{54}} < 0.$$

Since $c_{I,i} < c_{N,i}$, the result holds.

A.2 Proof of increases in $B_{N,i}$ increases propensity to insure

The individual insures if

$$U^I = U(\pi_i, c_{I,i}, X_i, B_i^{I*} - B_I, B_i^{I*} - B_{N,i}, \xi_i) > U(0, c_{N,i}, X_i, B_i^{N*} - B_I, B_i - B_{N,i}, \xi_i) = U^N.$$

An increase in $B_{N,i}$ will increase the propensity to insure if

$$U_4^I \frac{dB_i^{I*}}{dB_{N,i}} + U_5^I \left(\frac{dB_i^{I*}}{dB_{N,i}} - 1\right) - U_4^N \frac{dB_i^{N*}}{dB_{N,i}} + U_5^N \left(\frac{dB_i^{N*}}{dB_{N,i}} - 1\right) > 0.$$

From the Envelope Theorem,

$$U_4^I \frac{dB_i^{I*}}{dB_{N,i}} + U_5^I \frac{dB_i^{I*}}{dB_{N,i}} = U_4^N \frac{dB_i^{N*}}{dB_{N,i}} + U_5^N \frac{dB_i^{N*}}{dB_{N,i}} = 0.$$

Thus, I need only show that $U_5^N > U_5^I$. This result holds because from A.1 $B_i^{I*} > B_i^{N*}$.

A.3 Derivation of Multiselection Effects

Let X_i^I and X_i^C be respectively the observed variables that influence the decision to insure and the decision to visit a medical provide. The individual will insure if

$$X_i^I \beta_I + u_{i,I} > 0, \tag{15}$$

and will visit a provider if

$$X_i^C \beta_C + u_{i,C} > 0. \tag{16}$$

If the individual visits a provider then medical expenditures, C_i have gamma distribution with mean μ_i. I posit a log link function where

$$\begin{aligned} \ln \mu_i &= X_i^{C>0} \beta_{C>0} \\ &+E(\xi_i | \{X_i^C \beta_C + u_{i,C} > 0\} \\ &\cap \{X_i^I \beta_I + u_{i,I} > 0\} \end{aligned} \tag{17}$$

for insured patients and for uninsured patients

$$\begin{aligned} \ln \mu_i &= X_i^{C>0} \beta_{C>0} \\ &+E(\xi_i | \{X_i^C \beta_C + u_{i,C} > 0\} \\ &\cap \{X_i^I \beta_I + u_{i,I} \leq 0\}. \end{aligned} \tag{18}$$

I then posit

$$\begin{matrix} u_{i,I} \\ u_{i,C} \\ \xi_i \end{matrix} \, \tilde{} N(0, \Sigma). \tag{19}$$

Let $\Sigma_{i,j}$ be the (i, j) element of Σ. $\Sigma_{1,1} = \Sigma_{2,2} = 1$ and $\Sigma_{1,2} = \rho$. Then, from Manjunath and Stephan (2012)

$$\begin{aligned} E(\xi_i | \{a_1 &< u_{i,I} < b_1\} \cap \{a_2 < u_{i,C} < b_2\}) \\ &= \Sigma_{1,3} E_{1,i} + \Sigma_{2,3} E_{2,i} \end{aligned} \tag{20}$$

and

$$\begin{aligned} E_{1,i} &= E(u_{i,I} | \{a_1 < u_{i,I} < b_1\} \cap \{a_2 < u_{i,C} < b_2\}) \\ E_{2,i} &= E(u_{i,C} | \{a_1 < u_{i,I} < b_1\} \cap \{a_2 < u_{i,C} < b_2\}). \end{aligned} \tag{21}$$

More specifically, let $c = 1/\sqrt{1-\rho^2}$

$$\begin{aligned} E(u_{i,I} | \{a_1 &< u_{i,I} < b_1\} \cap \{a_2 < u_{i,C} < b_2\}) \\ &= \phi(a_1) \left[\Phi((b_2 - \rho a_1)c) - \Phi((a_2 - \rho a_1)c) \right] \\ &\quad -\phi(b_1) \left[\Phi((b_2 - \rho b_1)c) - \Phi((a_2 - \rho b_1)c) \right] \\ &\quad +\rho\phi(a_2) \left[\Phi((b_1 - \rho a_2)c) - \Phi((a_1 - \rho a_2)c) \right] \\ &\quad -\rho\phi(b_2) \left[\Phi((b_1 - \rho b_2)c) - \Phi((a_1 - \rho b_2)c) \right]. \end{aligned} \tag{22}$$

Likewise, let $p = \Pr\{a_1 < u_{i,I} < b_1\} \cap \{a_2 < u_{i,C} < b_2\}$, then

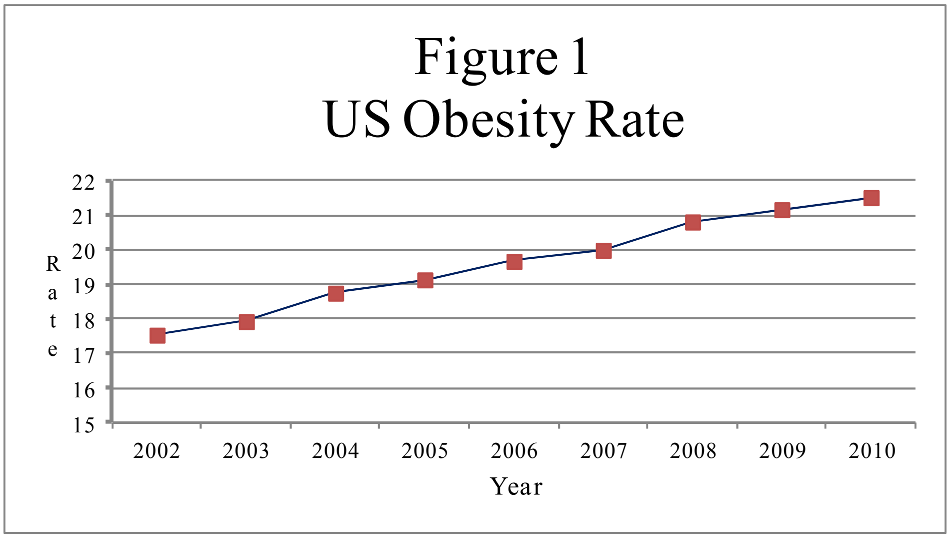
$$\begin{aligned} pE(u_{i,C}|\{a_1 &< u_{i,I} < b_1\} \cap \{a_2 < u_{i,C} < b_2\}) \\ &= \phi(a_2)\left[\Phi((b_1 - \rho a_2)c) - \Phi((a_1 - \rho a_2)c)\right] \\ &\quad -\phi(b_2)\left[\Phi((b_1 - \rho b_2)c) - \Phi((a_1 - \rho b_2)c)\right] \\ &\quad +\rho\phi(a_1)\left[\Phi((b_2 - \rho a_1)c) - \Phi((a_2 - \rho a_1)c)\right] \\ &\quad -\rho\phi(b_1)\left[\Phi((b_2 - \rho b_1)c) - \Phi((a_2 - \rho b_1)c)\right]. \end{aligned} \tag{23}$$

To estimate the selection effects, $\Sigma_{1,3}E_{1,i} + \Sigma_{2,3}E_{2,i}$, I start with a bivariate probit estimation of $I_i = 1$ and $C_i > 0$, or

$$\begin{aligned} \Pr(X_i^I\beta_I + u_{i,I} &> 0, X_i^C\beta_C + u_{i,C} > 0) \\ &= \Pr(-u_{i,I} < X_i^I\beta_I, -u_{i,C} < X_i^I\beta_I) \\ &= \Phi(X_i^I\beta_I, X_i^C\beta_C, \rho) \end{aligned} \tag{24}$$

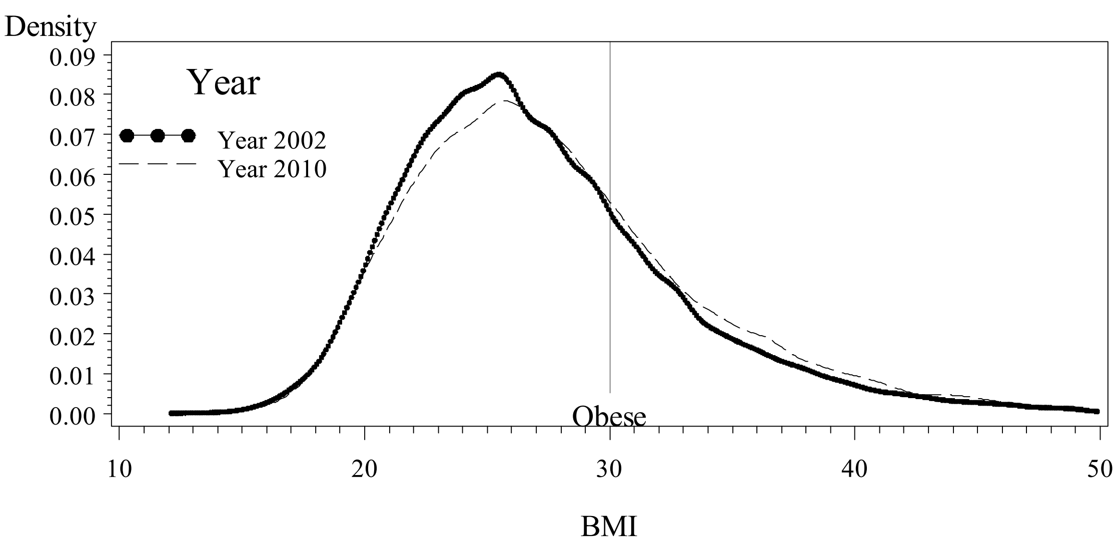
where $\Phi(.,.,.)$ is a standard bivariate normal distribution. Let $\widehat{\beta}_I, \widehat{\beta}_C, \widehat{\rho}$ be the parameter estimates from this bivariate probit estimation. Then if $I_i = 0$ and $C_i = 0$, I compute (23) and (24) by setting $a_1 = -\infty, b_1 = -X_i^I\widehat{\beta}_I, a_2 = -\infty$, $\beta_2 = -X_i^C\widehat{\beta}_C$, and $\rho = \widehat{\rho}$. Likewise if $I_i = 0$ and $C_i > 0$, then $X_i^I\beta_I + u_{i,I} \leq 0$ or $u_{i,I} \leq -X_i^I\beta_I$ and $X_i^C\beta_C > -u_{i,C}$. I compute (23) and (24) by setting $a_1 = -\infty, b_1 = -X_i^I\widehat{\beta}_I, a_2 = -\infty$, $\beta_2 = X_i^C\widehat{\beta}_C$, and $\rho = -\widehat{\rho}$. I do similar calculations for ($I_i = 1$ and $C_i = 0$) and ($I_i = 1$ and $C_i > 0$).

The parameters $\Sigma_{1,3}$ and $\Sigma_{2,3}$ are estimated as coefficients in the gamma regression of cost equation. Apparently, there is a negative coefficient for $\Sigma_{2,3}$. This is evidence that individuals with a high unobserved propensity not to see a provider (i.e. a highly negative $u_{i,C}$) will generate higher medical costs if they do see a provider because they have usually waited to long to see a provider and are sicker than they would have been if they had seen a provider sooner.



Source: Medical Expenditures Panel Survey

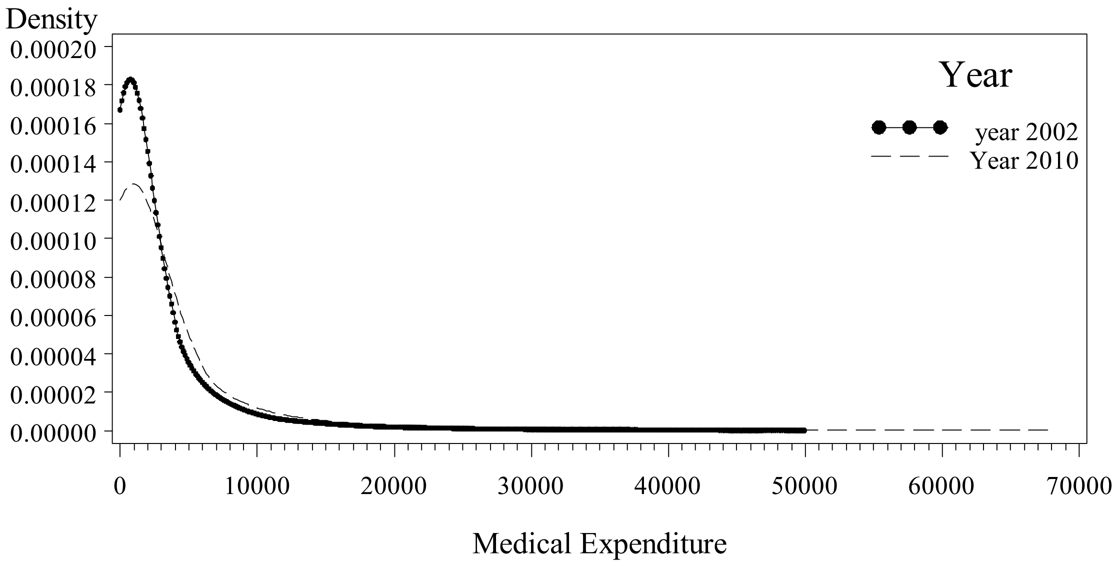
Figure 2
A Comparison of BMI Densities Between 2002 and 2010



Source: Medical Expenditures Panel Suvey 2002 and 2010

Figure 3

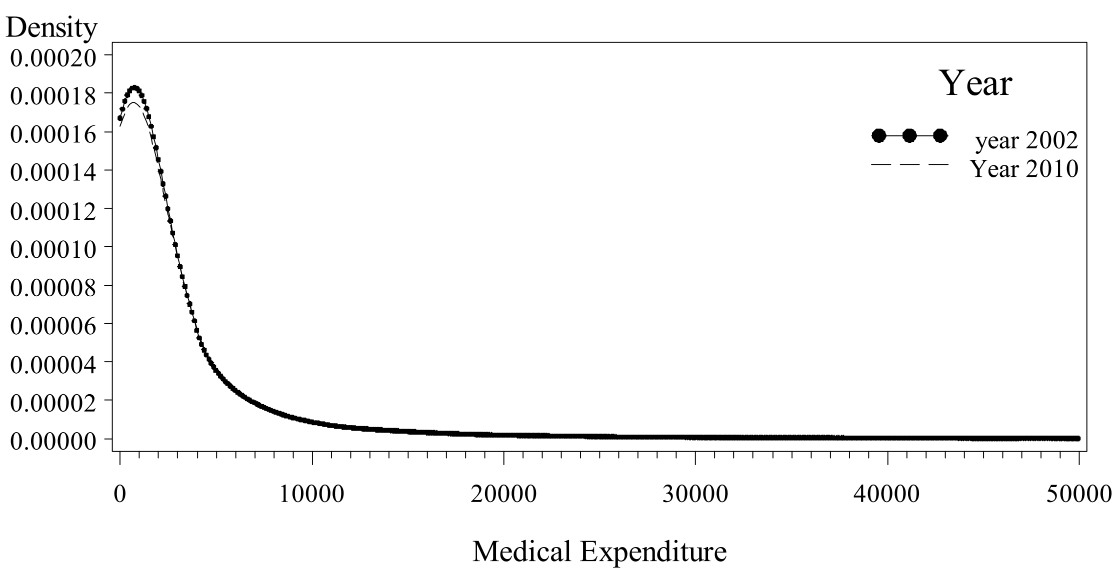
Per Person Nominal Medical Expenditure Densities Between 2002 and 2010



Source: Medical Expenditures Panel Suvey 2002 and 2010

Figure 4

Per Person Real Medical Expenditure Densities Between 2002 and 2010



Source: Medical Expenditures Panel Suvey 2002 and 2010

Table 1
Summary Statistics from the Medical Expenditures Panel Survey

Variable	Mean 2002 (Standard Deviation)	Mean 2010 (Standard Deviation)
Have a Usual Primary Provider	79.72% (0.44)	78.02% (0.49)
At least one zero cost visit	14.75% (0.32)	15.33% (0.34)
Do not see any provider	14.81% (0.32)	15.38% (0.34)
Black	12.32% (0.56)	12.49% (0.72)
Excellent Perceived Health	31.72% (0.46)	33.73% (0.53)
Male	48.86% (0.26)	49.12% (0.28)
Obese	17.54% (0.24)	21.53% (0.37)
Poor Perceived Health	2.86% (0.13)	2.78% (0.12)
Have Private Insurance	71.19% (0.62)	65.00% (0.77)
Have Public Insurance	17.05% (0.49)	21.89% (0.59)
Uninsured	11.75% (0.33)	13.10% (0.41)
Has Diabetes	4.84% (0.16)	6.81% (0.18)
Married	41.63% (0.41)	40.19% (0.49)
Student or Employed	52.78% (0.38)	51.72% (0.47)
Other Non Black Race	6.64% (0.37)	7.71% (0.58)
No Children	48.11% (0.52)	50.92% (0.66)
One Child	17.78% (0.38)	17.39% (0.46)
Two or More Children	34.11% (0.41)	31.69% (0.39)
Age	35.75	36.83

Variable	Mean 2002 (Standard Deviation)	Mean 2010 (Standard Deviation)
	(0.23)	(0.26)
BMI	26.99	27.75
	(0.05)	(0.06)
Years of Education	10.17	10.63
	(0.05)	(0.06)
Household Size	2.73	2.64
	(0.02)	(0.03)
Individual Income	22,166.50	25,711.01
	(270.38)	(365.05)
Per capita Expenditure in 2002 Dollars	2,813.24	3,010.42
	(59.12)	(68.26)
Imputed Per capita Expenditure in 2002 Dollars	3,406.23	3,498.54
	(74.39)	(79.33)
Nominal Per Capita Expenditure	2,813.24	4,094.38
	(59.12)	(92.84)
Imputed Nominal Per capita Expenditure	3,406.23	4,758.25
	74.39	107.89
Nominal per Capita Out of Poicket Payemnts	538.59	581.55
	(10.20)	(13.43)
Sample Size	37418	31228

Table 2
Estimates for
BMI Response Model

Variable	Estimate
Intercept	1.549**
	(0.095)
Male	0.327**
	(0.019)
Age	-0.004**
	(0.001)
Black	-0.014
	(0.032)
Employed or Student	0.099**
	(0.027)
Number of Children	0.090*
	(0.038)
Years of Education	0.026**
	(0.006)
Other Race	0.112**
	(0.042)
Household Size	-0.085*
	(0.036)

Variable	Estimate
Dummy for 2002	0.080*
	(0.035)
Dummy for 2003	0.086*
	(0.038)
Dummy for 2004	0.004
	(0.037)
Dummy for 2005	0.026
	(0.037)
Dummy for 2006	0.021
	(0.041)
Dummy for 2007	-0.033
	(0.037)
Dummy for 2008	0.076*
	0.033
Married	-0.051
	(0.040)
Household Income	0.000
	(0.000)
Individual Income	0.000
	(0.000)
Sum of Household's Years of Education	0.008*
	(0.004)
Number of High Occupations	-0.017
	(0.020)

Table 3
Control Equation for BMI

Variable	Estimate
Intercept	20.109**
	(0.803)
Age	0.009**
	(0.003)
Years of Education	0.106**
	(0.021)
Have a Provider	1.015**
	(0.039)
Price of Corn Syrup	-0.887
	(0.491)
Has Arthritis	-0.646**
	(0.026)
Black	1.673**
	(0.054)
Spouses Income	0.000**
	(0.000)
Employed or Student	0.928**
	(0.077)
Household Size	-0.106**
	(0.045)
Male	3.464**

Variable	Estimate
	(0.186)
Number of High Occupations	-0.146**
	(0.030)
Number of Children	0.334**
	(0.054)
Other Race	-0.974**
	(0.085)
Dummy for 2002	-0.508**
	(0.151)
Dummy for 2003	-0.384**
	(0.137)
Dummy for 2004	-0.826**
	(0.137)
Dummy for2005	-0.609**
	(0.143)
Dummy for2006	-0.478**
	(0.117)
Dummy for2007	-0.742**
	(0.078)
Dummy for2008	0.432**
	(0.072)
Inverse Mill	103.333**
	(7.416)
Inverse Mill Sq	-275.670**
	(22.837)

Table 4
Insurance Selection with BMI as Endogenous

Variable	Estimate
Intercept	-2.46112**
	(0.1131)
Age	-0.00009
	(0.0016)
Years of Education	0.11093**
	0.00488
BMI	0.00510**
	(0.0012)
Male	-0.39402**
	(0.0803)
Age*male	0.00464**
	(0.0018)
Household Size	0.21084**
	(0.0375)
Individual Income	0.00002**
	(0.0000)
Total Household Income	0.00001**
	(0.0000)
Number of High Occupations	0.58573**
	(0.0319)
Black	0.10209*
	(0.0401)
Perceived Poor Health	-0.23748**
	(0.0556)
SCHIP Children	-0.86159**
	(0.0300)
Perceived Excellent Health	0.06659*
	(0.0286)
Number of Children	0.00612
	(0.0517)
Have a Primary Provider	1.21294**
	(0.0273)
Dummy for2002	0.37017**
	(0.0493)
Dummy for2003	0.29947**
	(0.0489)
Dummy for2004	0.26344**
	(0.0497)
Dummy for2005	0.18161**
	(0.0499)
Dummy for2006	0.09834*
	(0.0466)
Dummy for2007	0.09592*
	(0.0452)
Dummy for2008	0.02465
	(0.0365)

Table 5
BMI Choice with Insurance as Endogenous

Variable	Estimate
Intercept	19.333**
	(0.380)
Age	0.015**
	(0.002)
Years of Education	0.106**
	(0.013)
Insidividual Income	0.000**
	(0.000)
Privately Insured	1.528**
	(0.075)
Male	2.805**
	(0.110)
Number of Children	0.631**
	(0.045)
Black	1.661**
	(0.048)
Other Race	-1.322**
	(0.069)
Household Size	-0.390**
	(0.036)
Employed or Student	0.711**
	(0.055)
Response Mill	87.090**
	(4.326)
Response Mill Squared	-206.489**
	14.540
Insurance Mill	-1.314**
	(0.072)

Table 6
Parameter Estimates of
Multivariate Probit Model

Insurance Propensity	Estimate
Intercept	-1.545**
	(0.037)
Age	0.002**
	(0.001)
Years of Education	0.064**
	(0.002)
BMI	0.002**

Propensity to Visit Provider	Estimate
Intercept	-3.240**
	(0.277)
Male	-0.539**
	(0.013)
Age	0.010**
	(0.001)
Black	-0.327**

Insurance Propensity	Estimate
	(0.000)
Male	-0.106**
	(0.036)
Age*Male	0.001
	(0.001)
Household Size	0.100**
	(0.011)
Individual Income	0.873**
	(0.034)
Sum of Family Income	0.644**
	(0.026)
Number in high Occupations	0.346**
	(0.009)
Black	0.112**
	(0.013)
Poor Perceived Health	-0.133**
	(0.027)
SCHIP Household	-0.438**
	(0.007)
Perceived Excellent Health	0.081**
	(0.011)
Number of Children	0.028*
	(0.013)
Have Primary Provider	0.710**
	(0.010)
2002 Dummy	0.223**
	(0.016)
2003 Dummy	0.163**
	(0.017)
2004 Dummy	0.139**
	(0.017)
2005 Dummy	0.096**
	(0.017)
2006 Dummy	0.062**
	(0.017)
2007 Dummy	0.067**
	(0.017)
2008 Dummy	0.034*
	(0.017)

Propensity to Visit Provider	Estimate
	(0.022)
Poor Perceived Health	0.710**
	(0.040)
BMI	0.090**
	(0.010)
Employed or Student	-0.063**
	(0.013)
Number of Children	0.052**
	(0.018)
Years Education	0.054**
	(0.003)
Other Race	-0.067**
	(0.024)
Household Size	-0.127**
	(0.017)
2002 Dummy	0.174**
	(0.018)
2003 Dummy	0.174**
	(0.019)
2004 Dummy	0.106**
	(0.017)
2005 Dummy	0.125**
	(0.017)
2006 Dummy	0.096**
	(0.017)
2007 Dummy	0.111**
	(0.017)
2008 Dummy	0.047**
	(0.016)
Married	0.009
	(0.019)
All Income	0.284**
	(0.025)
Individual Income	0.206**
	(0.031)
Sum of Households Education Years	0.012**
	(0.002)
Number in High Occupations	0.149**
	(0.009)

Insurance Propensity	**Estimate**

Propensity to Visit Provider	**Estimate**
Have Primary Provider	0.816**
	(0.014)
ρ	0.310**
	(0.006)

Table 7
Parameter Estimates for
Cost Equation

Variable	Cost Not Imputed	Cost Imputed
Intercept	7.706**	7.606**
	(0.039)	(0.040)
BMI	0.003**	0.003**
	(0.001)	(0.001)
Male	-0.117**	-0.184**
	(0.011)	(0.011)
Perceived Poor Health	1.002**	1.114**
	(0.030)	(0.031)
Have a primary provider	0.071**	0.211**
	(0.020)	(0.020)
Perceived Excellent Health	-0.421**	-0.444**
	(0.010)	(0.011)
Black	-0.001	0.020
	(0.016)	(0.016)
Age	0.018**	0.019**
	(0.000)	(0.000)
Employed or Student	-0.300**	-0.298**
	(0.013)	(0.013)
Other Race	-0.042*	-0.112**
	(0.018)	(0.019)
2002 Dummy	-0.146**	-0.112**
	(0.017)	(0.017)
2003 Dummy	0.006	0.027
	(0.017)	(0.017)
2004 Dummy	-0.007	0.013
	(0.017)	(0.017)
2005 Dummy	-0.015	0.023
	(0.016)	(0.017)
2006 Dummy	-0.061**	-0.030
	(0.016)	(0.017)
2007 Dummy	0.009	-0.003
	(0.016)	(0.017)
2008 Dummy	-0.058**	-0.033
	0.016	0.017
E1	0.227**	0.191**

Variable	Cost Not Imputed	Cost Imputed
	(0.009)	(0.010)
E2	-0.928**	-0.372**
	(0.040)	(0.041)

Table 8
Impact of Obesity on Medical Cost

Cost of Obesity	**Estimate**
Average Direct Cost of Obesity	$430.52
Cost from Additional Propensity to Insure	$3.83
Cost Reduction from Increase Propensity to Visit Provider	-$4.02
Cumulative Effects	$430.33

Average Effect of a 10% Reduction in BMI	**Estimate**
Average Direct Effect	-$45.44
Effect from Reduced Propensity to Insure	-$0.63
Effect from Reduced Propensity to Visit Provider	$0.79
Cumulative Effects	-$45.28

%Reduction in Cost from a %1 Reduction in BMI	**Estimate**
Direct	0.0115%
Effect from Insurance Propensity	0.0002%
Effect from Visiting Propensity	-0.0003%
Cumulative Effects	0.0115%