Using Social Simulation to Understand the Obesity Epidemic in the United States
Michelle Saksena

Abstract:

Despite a growing concern for the obesity epidemic in the United States, data constraints have limited social science analyses to correlative studies. This limits our capacity to make causal statements about which factors contribute to weight gain. This is further compounded by the fact that the nature of the obesity problem is plagued with several endogeneity problems. Currently, no national database within the United States concurrently collects longitudinal data on individuals’ weight outcomes, dietary intake and economic and geographic indicators. However, this form of data is necessitated by traditional econometric approaches in order to treat endogeneity and thus has thwarted the advancement of the study of obesity as a social problem.

For this reason, I approach the obesity epidemic as a complex system. Wherein by, endogeneity is not only integrated into the model, but is regarded as an autonomous factor contributing to micro and macroscopic outcomes of the system. In this sense I have characterized the obesity problem as an adaptive system where people’s actions and preferences are governed by their local environments, history of food choices and consequently, their changed weight status as a result of those choices over time. I am particularly interested in the consumptive behavior of individuals by income class. I theorize that the lowest income individuals will tend to be the most obese because they are relegated to food choices that are calorically dense, but nutritionally lax and thus will tend to over-consume.

Results show that in fact the lowest income individuals do not tend to be the most overweight. Rather, it is the middle class that exhibits the highest preponderance of obesity. This suggests that poor individuals are still dominated by a large negative income effect and simply do not possess enough purchasing power to consume a sufficient number of calories. While, middle class individuals—despite the ability to purchase an abundance of calories—may still be out priced by health and may over consume to compensate for nutritional deficiencies. When a tax is applied there is a shift of burden of obesity to the upper middle class since real income is deflated and as such middle class individuals lose some purchasing power which would otherwise enable them to increase their obesity status.
Obesity has become a widespread social problem. While the majority of research in obesity focuses on isolating static, microscopic factors leading to individual-level obesity, limited work has been done to treat obesity as an adaptive, public health issue. Despite the shortage of research using this framework, the issue of obesity would benefit from a systemic analysis as the repercussions are not merely relegated to individuals, but collectively burden healthcare systems and have adverse effects on labor markets.

In the United States, healthy foods cost more than their less nutritious counterparts, which are typically calorically dense yet devoid of nutrients. It is therefore reasonable to hypothesize that poorer individuals will have a higher propensity to be overweight since they maybe out priced by health, but have access to an abundance of low quality calories. Furthermore, overconsumption by poorer individuals may in fact be a rational response to nutritional deficiencies in their diets despite leading to weight gain.

There is a presiding belief that low income is associated with inferior dietary quality in developed nations (Sobal and Stunkard 1989; Lakdawalla and Philipson 2002; Drewnowski and Darmon 2005; Thomas and Frankenberg 2002; Schmeiser 2008). Though admittedly in the minority, more recent studies refute or indicate little relationship with obesity and low income in developed nations. Cawley et al. (2010) found no significant impact of income on weight among the elderly in the United States while Griffith et al. (2013) found that UK middle class children showed highest prevalence of obesity. Griffith et al. (2013) goes on to state that perhaps the relationship between income and obesity is not linear, but rather is quadratic with income. Prentice’s (2005) article is perhaps the most alarming, as he cautions that within the near future obesity will no longer be confined to developed nations, but will also infiltrate poorer countries. This of course contradicts Prentice’s antecedents, but he denotes that empirical findings are
subject to rapid obsolescence due to the precipitous expansion of the epidemic. This very fact compels for prompt and effective public intervention.

The purpose of this paper is to study the effects of income on the proliferation of obesity among populations. This paper utilizes simulation to address the relationship between income and obesity and whether consumptive behavior changes as a result of an individual’s social environment. In this case simulation methods are an appropriate treatment of the issue for three reasons. First, as mentioned previously obesity is spreading at an alarming rate making empirical findings quickly out of date. Therefore, simulation is a platform that can be used to envisage many future scenarios in a much smaller time frame than that of real life. Data availability has also limited empirical studies, as currently there does not exist a national dataset, which is sufficiently longitudinal and collects information on income, geographic information, demographics, dietary behavior and obesity outcomes. Second, the obesity epidemic possesses many of the characteristics of a complex system (Hammond 2009). Particularly, the pathways in which an individual becomes obese are numerous and are often studied across many disciplines. This high level of heterogeneity among agents is further compounded by an equally varied set of environmental factors. Finally, interactions between agents and their environments are complex in that exchanges are often endogenous. In fact, the existence of feedback loops in and of themselves maybe an attributing factor to obesity. However, these non-linear, cyclical interactions make it quite challenging to study obesity particularly in an empirical setting and as a result are not well understood.

Most articles that tie the application of complex systems to the obesity epidemic have largely been expositional in nature and do not present a formal structural model (Hammond 2009; Prince 2009; Bruzzone, Novak and Madeo 2012). The exception is Auchinclose et al.
(2011) whereby they present an agent based model, which looks at how the presence of food deserts in poor communities may attribute to higher preponderance of obesity because of lack of access to healthy foods. This paper is congruent with Auchinclose et al.’s (2011) etiology of the formation of local environments and more specifically, mobility constraints. A Tiebout (1956) sorting model is used to geographically distribute individuals and subsequently the creation of income based neighborhoods. However, beyond this application, the issue of “access” is deemphasized. Rather my perspective is more cynical in that I believe that individual behavior will not merely change simply if there is increased access to healthy foods. More likely, individuals will exhibit consumption behavior that optimizes their recursive utility maximization problem, which may not necessarily preclude them from unhealthy choices. By subjecting agents to different price environments, the results may become an informative resource to constructing effective tax policy to curb obesity.

A deterministic model was first run and helped to illuminate the growth rate of BMI over time. Because of a lagged metabolic response to caloric intake it was shown that unlike conventional wisdom which states that weight gain increases linearly over time, BMI actually increased logarithmically. This relationship insinuates that short term response to high caloric intake has much larger effects on body weight, but tapers off over time as the individual acclimatizes to higher food intake.

Results of the baseline model are encouraging. After 1737 Monte Carlo simulations the averages and standard deviations were consistent with the BMI distribution from the National Health and Nutrition Survey 2009-2010 (NHANES) dataset. The average mortality rate

---

1 In this model I considered mortality to be individuals with a BMI < 13.18, which is the minimum BMI value found in the NHANES 2009-2010 dataset. The rationale behind this was that the NHANES data is interpreted to be naturally truncated with only living persons. I assume then that 13.18 is statistically the minimal BMI level for an individual to be alive. However, actual biological limitations may be different.
attributed to malnutrition was much higher in the simulation than what is quoted by the World Health Organization (WHO), which calculates a mortality rate of about 0.045% in the United States. However, this rate is within the range of mortality rates exhibited by the 1737 simulations. Implementing a menial tax\(^2\) of $0.04 on carbohydrates shows a dramatic effect on the BMI population average. Interestingly, incidence of obesity shifted from middle class to upper middle class or households within the 75\(^{th}\) and 50\(^{th}\) percentiles as relative purchasing power decreased in the taxed environment. Mortality rates also increased in the more expensive price environment. In reality the presence of food assistance programs such as WIC or SNAP may play a large part in alleviating death caused by malnutrition and maybe the reason for the mortality rate discrepancy.

The first section presents a theoretical model. The second section describes the computational model using the standardized ODD protocols (Grimm et al 2006; Grimm et al 2010; Rand and Rust 2011). The third section discusses results of Monte Carlo simulations by varying initial conditions. The fourth section highlights the limitations and scope of this simulation. Finally, the fifth section contains conclusions.

**Theoretical Framework**

1.1 Ex-Ante Formation of Neighborhoods

There is much discussion surrounding the implications of limited access to healthful foods and its effects of obesity among the poor. Though the concept of food deserts is not new to obesity studies it is deemphasized in this study in favor of focusing on price effects. However, it is would be unreasonable to assume that each individual is faced with equally randomized

---

\(^2\) I used pricing data from Brooks et al. (2010), which claim the price of carbohydrates in megajoules (MJ) per gram to be $-0.38. However, since I used a cost minimization problem to calculate the quantities of each macronutrient in each meal it was impossible to use this price as it is negative. As a result I adjusted the price for the baseline model to be $0.01 and then compensated household income for the increase in price. All income levels are then interpreted real income and represent the real purchasing power of each individual in each simulation and price environment.
environmental effects. For this reason I use the Tiebout (1956) sorting model to form neighborhoods of like income households. The rationale behind this is as Tiebout (1956) illuminates; households will locate to communities with like attributes of their household. In this model I use income to distribute the geography of households. Households with similar income ranges will inevitability face similar, but not identical environmental constraints. Figure 1 is a picture of one simulation of a Tiebout Sorting model. The result is the formation of poor, middle class and rich neighborhoods where the buildings represent restaurants and the individuals represent the location of their household.

Figure 1.

1.1 Utility Maximization Problem

In this model individuals maximize their utility by consuming calories, but are also conflicted by their preferences to maintain weight. As such, an individual’s biological desire to consume a maximal amount of calories at each meal is diminished by her desire to also stay healthy. And, these preferences of satiety versus health differ among individuals and over time.
Comparably, Richards and Patterson (2006) assume that ex-ante, each individual decides his or her idealized nutrient profile. In their model a consumer’s cost minimization problem is subject to some technology of producing nutrients from food while a law of motion equation, defines the depreciation of the individual’s nutrient stock over time.

\[ \max_{\text{satiety}_{i,t}, \text{health}_{i,t}} U_{i,t} = \text{satiety}_{i,t}(\text{BMR}_{i,t-1}, \text{kcals}_j) \alpha \text{health}_{i,t}(\text{BMR}_{i,t-1}, \text{kcals}_j)^{1-\alpha} + \delta U_{i,t-1}(\text{BMR}_{i,t-2}, \text{kcals}_j) \]  

(1)

where,

\[ 0 < \alpha_i(\text{obesity}_{i,t-1}), \delta < 1 \] \text{ and } \[ \alpha_i'(\text{obesity}_{i,t-1}) < 0 \]

s.t.

\[ \text{income}_i \leq p_{fr}pr + p_{ff} + p_ec \]
\[ \text{travel radius}_i \leq TCF_i * \text{distance}_j \]

This model characterizes individual behavior as a utility maximization problem over time and is shown in equation (1). In this model, individuals derive utility from calories and weight status only. Each individual chooses between two goods: satiety and health. There is no differentiation between the originations of calories. Accordingly, a calorie from fat is no different from a calorie from protein or carbohydrate. Rather than having people exogenously decide their preferred nutrient profiles each individual is assigned an initial body mass index (BMI) a priori from a distribution calibrated to data using maximum likelihood methods. A person’s preferences for satiety and health are informed by their weight status. Since overall an individual desires to maintain his or her weight it is assumed that as weight status increases an individual will prefer health over satiety and thus \( \alpha_{i,t} \) is decreasing in weight status, \( \text{obesity}_{i,t} \).

Utility in the current period is also a function of a previous utility as it is reasonable to assume

\[ \text{A nutrient profile is considered to be the combination of the three macronutrients: carbohydrates, fat and protein and each person’s preferred nutrient profile is a reflection of his or her preferences for satiety and health. The total amount of calories is calculated as a linear combination of all three nutrients.} \]

\[ \text{Stigler and Becker 1977, Becker and Murphy 1988 and Iannaccone 1986 previously mention the idea of depreciation of nutrient stocks over time. Richards and Patterson (2006) incorporate this into their model.} \]

\[ \text{National Health and Nutrition Examination Survey (NHANES) 1999-2000 dataset} \]
that utility from previous meals influence the choices of an individual in the current period.

Each individual is subjected to prices of each macronutrient \( p_{pr}, p_c, p_f, p_{tg} \), which were taken from Brooks et al.’s (2010)\(^7\) empirical findings. The total price of a meal was calculated as the composition of protein, carbohydrate and fat as well as the total grams of the meal times their prices. Since the price of protein was highest, meals that were protein rich tended to be the most expensive.

Each individual was also subjected to a mobility constraint. Mobility was decided as a function of income class. Poorer individuals are most restricted since it is reasonable to assume that access to transportation is also a function of income hence their travel cost factor (TCF) should be higher than their richer counterparts. The mobility constraint essentially creates a travel radius for each individual and is centered around each person’s household location.

1.1.1 Satiety

Equation (2) illustrates the functional form of satiety. Satiety is modeled as a sinodial curve with dampening amplitude. Satiety is naturally bounded by \([0, 3420]\)\(^8\) as no meals in the simulation are below zero or above 3420 calories. The functional form is such that the argmax is exactly one third of the individual’s Basal Metabolic Rate (BMR) or the daily amount of calories

\[
Satiety_{i,t} = \frac{BMR_{i,t-1}}{3} e^{\frac{-1}{BMR_{i,t-1}+kcal_{j}}} \sin \left( \frac{1}{4 \times BMR_{i,t-1}} \pi \frac{kcals}{5} \right)
\]

\( p_{pr} = 3.26, p_c = -0.38(0.01), p_f = 0.086. \) Prices are in megajoules (MJ) per gram of corresponding macronutrient.

\( p_{tg} \) is the price of protein, which was taken from Brooks et al.’s (2010)\(^7\) empirical findings. The total price of a meal was calculated as the composition of protein, carbohydrate and fat as well as the total grams of the meal times their prices. Since the price of protein was highest, meals that were protein rich tended to be the most expensive.

Equation (2) illustrates the functional form of satiety. Satiety is modeled as a sinodial curve with dampening amplitude. Satiety is naturally bounded by \([0, 3420]\)\(^8\) as no meals in the simulation are below zero or above 3420 calories. The functional form is such that the argmax is exactly one third of the individual’s Basal Metabolic Rate (BMR) or the daily amount of calories

\( Satiety_{i,t} = \frac{BMR_{i,t-1}}{3} e^{\frac{-1}{BMR_{i,t-1}+kcal_{j}}} \sin \left( \frac{1}{4 \times BMR_{i,t-1}} \pi \frac{kcals}{5} \right) \)

\( p_{pr} = 3.26, p_c = -0.38(0.01), p_f = 0.086. \) Prices are in megajoules (MJ) per gram of corresponding macronutrient.

\( p_{tg} \) is the price of protein, which was taken from Brooks et al.’s (2010)\(^7\) empirical findings. The total price of a meal was calculated as the composition of protein, carbohydrate and fat as well as the total grams of the meal times their prices. Since the price of protein was highest, meals that were protein rich tended to be the most expensive.

\( p_{tg} \) is the price of protein, which was taken from Brooks et al.’s (2010)\(^7\) empirical findings. The total price of a meal was calculated as the composition of protein, carbohydrate and fat as well as the total grams of the meal times their prices. Since the price of protein was highest, meals that were protein rich tended to be the most expensive.

\( p_{tg} \) is the price of protein, which was taken from Brooks et al.’s (2010)\(^7\) empirical findings. The total price of a meal was calculated as the composition of protein, carbohydrate and fat as well as the total grams of the meal times their prices. Since the price of protein was highest, meals that were protein rich tended to be the most expensive.
burned by the individual at time $t$.

Figure 2.

![Satiety Curve](image)

Figure 2 is a depiction of a person’s satiety curve whose BMR is precisely 1500 calories. The first segment of the curve can loosely be interpreted as a Laffer curve. The more calories an individual consumes the more utility he receives up until a global maximum where $\text{argmax} \equiv \text{BMR}/3$. Beyond this point the individual receives decreasing levels of utility, as he would prefer to maintain his weight. Utility from excess calories further decreases to the point of negative utility levels, however at some point utility levels start to increase again. The reason behind this is that at some point, excess calories are so much larger than what is ideal that the individual changes his preferences in favor of pure satiety. The minimum point can be thought of as the “point of resignation” and consumption of calories at or beyond this point will increase utility though the maximal amount will not exceed the utility derived from eating a third of ones BMR.

1.1.2 Health

$$\text{health}_{i,t} = \frac{1}{\text{kcals}_{j} - \frac{\text{BMR}_{i,t-1}}{3}}$$

The functional form of health is much simpler and is measured as the inverse of the
distance between calories consumed from the meal and a third of the individuals BMR at t. Deciding to represent health in a rather simplistic way was primarily due to a lack of scientific consensus as to which foods are actually considered healthy. This functional form is intentionally an innocuous representation of health as the practice of deeming foods as healthy is more exclusionary and in many ways is reflective of diet trends rather than by science. Furthermore, it would be dubious to implicate “healthiness” on individual foods, as health should be a measure of average diet consumption rather than on a particular food. While it maybe safe to say that foods that are highly processed, laden with sugar and fat are unhealthy, classifying foods that do not advertently fall under all of these categorizes is still quite subjective.

Protein consumption maybe a fitting measure of diet quality as Richards and Patterson (2006) found that higher protein consumption was associated with lower caloric consumption overall among Native Americans. However, protein consumption was interpreted in aggregate consumption levels and is not a suitable measure in this particular case. Also, since I use exogenous prices for each macronutrient, quality should be reflected in the prices. Heuristically, this appears to be the case as the price of protein is much higher than carbohydrates and fat.

1.2 Cost Minimization Problem

To determine the caloric content of meals restaurants and households are faced with a cost minimization problem subjected to \( p_{pr} \), \( p_c \) and \( p_f \). This problem regards prices from Brooks et al. (2010) as input prices.\(^9\) While the production function of each household and restaurant corresponds to the creation of meals. The cost minimization problem and input factor demand functions for protein, carbohydrate and fat are as follows.

\(^9\) The prices from Brooks et al. (2010) were calculated from retail prices of grocery foods. This roughly translates to specific macronutrient output prices. However, I regard the household and the restaurant as producers of meals and not of individual foods. Therefore, it is justifiable to interpret these prices as prices for input materials (grocery ready foods) that go into preparing meals.
\[
\min_{c, pr, f} C = p_c \cdot c + p_{pr} \cdot pr + p_f \cdot f
\]  \quad (4)

s.t.
\[
\bar{Q}_j = F(c, pr, f) = Ac^\alpha pr^\beta f^\gamma
\]

where,

\[ A \equiv 1 \text{ and } \alpha + \beta + \gamma = 1 \]

The optimal input factor demands:
\[
c^* = \frac{\bar{Q}_j}{\left(\frac{\beta}{\alpha} + \frac{p_c}{p_{pr}}\right)^\beta} \times \left(\frac{\gamma \cdot p_c}{p_f}\right)^\gamma
\]  \quad (5)

\[
pr^* = \frac{p_c}{p_{pr}} \times \frac{\beta}{\alpha} \times c^*
\]  \quad (6)

\[
f^* = \frac{p_c}{p_f} \times \frac{\gamma}{\alpha} \times c^*
\]  \quad (7)

1. Computational Model

2.1 Purpose

This model was designed to explore questions about obesity outcomes. In this model I am interested in unraveling obesity outcomes of populations under different price environments.

2.2 Entities, State and Calculated Variables and Scale
The model has four kinds of entities: individuals, households, restaurants and patches of land. The patches make up a square grid landscape that is $25 \times 25$. Each patch is categorized by type. Type 0 or null type is designated as an empty patch, type 1 is a residential patch and type 2 is a commercial patch. Households can only exist on type 1 patches while restaurants may only occupy type 2 patches. Each household and restaurant offers one representative meal, which is characterized by the following state variables: output $\bar{Q}$, prices $p_{t, g}, p_{pr}, p_f, p_c$ and output elasticities $\alpha, \beta, \gamma$. $\bar{Q}$ is determined from random draws of a normal distribution with $\mu=15$, $\sigma=6$. Input prices for total grams and each macronutrient are determined empirically (Brooks et al. 2010). Output elasticities are taken from normalized grams to calories conversion rates for each macronutrient. Each gram of carbohydrate and protein is equivalent to four calories while one gram of fat is equal to nine calories.

Levels of carbohydrates, fat and protein are in grams and each are calculated as the optimal level of their input factor demands given by equations (5), (6) and (7). Total grams is calculated post optimization and is equal to the sum of carbohydrates, protein and fat. Total calories is equal to each macronutrient in grams multiplied by the appropriate gram to calorie conversion rate. Price is calculated similarly as the sum of each macronutrient times the price while price class is determined by the household or restaurant’s percentile score.
Individuals are characterized by many state variables. Each individual is assigned a sex, age, body mass index (BMI), height, income and leisure hours. Sex is determined by a binomial distribution. Age is bounded between 18 and 80 and draws are from a discrete uniform distribution. Using the National Health and Nutrition Examination Survey (NHANES) 1999-2000 dataset I use maximum likelihood methods to fit a maximum extreme value distribution ($\lambda = 20, \kappa = 5$) to BMI data. Starting BMI levels were drawn from this distribution. Height was

\[ \text{BMI}_{\text{age}} \times \text{height}^2 \]

\[
\begin{align*}
\text{BMI}_{\text{age}} &= 1, \text{ if income}_i \in \text{first tertile} \\
&= 2, \text{ if income}_i \in \text{second tertile} \\
&= 3, \text{ if income}_i \in \text{third tertile} \\
\end{align*}
\]

\[
\begin{align*}
\text{income class}_i &= 1, \text{ if income class}_i = 1 \\
&= 2, \text{ if income class}_i = 2 \\
&= 3, \text{ if income class}_i = 3 \\
\end{align*}
\]

\[ f(\text{price}_i) \]

\[ \text{Also known as a Gumbel Distribution} \]
determined from a uniform distribution and is bounded by 152 to 183 centimeters. Leisure hours are determined from a discrete uniform distribution bounded by 5 and 20 hours.

Lastly, a two-step process determines income. First, price\textsubscript{j} from households and restaurants are used as base income for each individual. In order to rid rounding errors\textsuperscript{11}, income is initially calculated to be 1.05*price\textsubscript{j} (in this case price\textsubscript{j} corresponds to the price of a meal offered at the individual’s household). This alone caused uncharacteristically\textsuperscript{12} high rates of BMI attrition\textsuperscript{13} over the course of the simulation. So, a subsidy\textsuperscript{14} was created for households in the lowest third of the income distribution. The poorest households were allocated an income equal to the income level of the 16.67\textsuperscript{th} percentile household or the first 1/6\textsuperscript{th} of the income distribution. This income was multiplied by 5 and is the new minimum income a household can have in this simulation. Households who are at or below the 33\textsuperscript{rd} percentile, but above the first 1/6 of the income distribution also receive the same income subsidy up till the 16.67\textsuperscript{th} percentile income level.

The timeline for each simulation runs over a representative ten years. Each time increment (tick) is interpreted as one meal event and three consecutive ticks are considered to be one day. Therefore, the simulation runs for 10950 ticks\textsuperscript{15}.

\textsuperscript{11} This step was necessary in order to ensure that every individual could at least afford to consume at least one meal, which was by default the meal available at one’s household. This was of particular importance to the poorest individuals as they also have more mobility constrains and may not necessarily be able to patronize a restaurant which offers a cheaper meal if it is outside of their travel radius.

\textsuperscript{12} According to 2004 numbers, WHO records deaths attributable to malnutrition in the United States to be roughly 45 deaths per 100,000 individuals.

\textsuperscript{13} BMI attrition is defined as the percentage of individuals whose BMI fell below 13.18, or the statistical minimum value found in the NHANES 2009-2010 data set.

\textsuperscript{14} In absence of the subsidy this model most closely represents free markets for food consumption. In reality, many foods themselves are highly subsidized, while the presence of social programs helps to elevate food insecurity among the poorest. So, the purpose of adding a subsidy in this context is to emulate actual purchasing power of the poorest households.

\textsuperscript{15} (3 * 365 * 10) = 10950
2.3 Process Overview and Scheduling

Prior to the optimization, each household and restaurant must migrate to the appropriate neighborhood. The order in which agents move is irrelevant during this step as the process assumes that agents move simultaneously. The aggregate result is the formation of neighborhoods by income.

During the optimization only individuals move. Each individual selects his optimal choice and moves to that destination. After each individual moves to his optimal destination he “consumes” the meal and moves back to his household. Each meal has a caloric content, which is inherited by the individual and incrementally changes the person’s weight. Here BMR also adjusts to accommodate the caloric intake. Over time each person’s weight changes as a consequence of his or her historical caloric intake and as a result, his or her preferences for satiety and health also change over time. Again, the succession of individuals is unimportant since it is assumed that all individuals consume meals simultaneously at each time increment while the scarcity assumption\(^\text{16}\) is relaxed.

2.4 Initialization

In each simulation there are 300 individuals and households as well as an additional 201 restaurants. As state earlier, neighborhoods are formed using a Tiebout (1956) sorting model, which establishes the locations of restaurants and households prior to the optimization simulation. Every individual, household and restaurant is characterized by their initial state variables explained in Tables 1 and 2.

\(^{16}\) It is assumed that each restaurant/household can supply infinite number of meals. Therefore, there are no restrictions on the number of patrons per restaurant/household at any time.
2.5 Input Data

The only sources of input data come from macronutrient prices taken from Brooks et al. (2010) and NHANES 1999-2000 data used to formulate the distribution of initial BMI’s.

2.6 Submodels

2.6.1 Tiebout Sorting Model

First restaurants and households are categorized by income/price level. The poorest households and restaurants correspond to the first tertile of the income/price distribution while middle and rich classes correspond to the second and third tertile respectively. Initially, households and restaurants are distributed onto the landscape randomly. Each restaurant/household randomly moves to a vacant patch if less than 66% of neighboring patches are not also occupied by like households or restaurants. The sorting model converges when all agents have stopped moving. Once sorting has completed each restaurant and household inherit the patch’s coordinates as its permanent location.

2.6.2 Optimization

<table>
<thead>
<tr>
<th>빈도</th>
<th>빈도</th>
<th>빈도</th>
<th>빈도</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

12 Each BMI range and obesity classification was established by the World Health Organization (WHO) (“Redefining Obesity” 2000).
Once the neighborhoods are formed the actual optimization occurs. All restaurants within the person’s travel radius and his or her own household are included as possible destinations. Then, restaurants whose meal price exceeds the individual’s income constraints are excluded from possible destinations. This forms the individual’s feasible choice set, $\mathcal{F}_{i,t}$. For each destination a utility is calculated at time $t$ using equation (1). Individual’s preferences for satiety and health are determined by their current BMI. Table 3. shows the corresponding $\alpha$ by BMI status. Once utilities are calculated the individuals move to the destination that offers the highest level of utility.

2.6.4 Update Variables

Females: $BMR = 88.362 \times 4.799 \times \text{height}(m^2) + 13.397 \times \text{weight}(kg) - 5.677 \times \text{(age)} \times \text{activity level}$

Males: $BMR = 447.593 \times 3.098 \times \text{height}(m^2) + 9.247 \times \text{weight}(kg) - 4.330 \times \text{(age)} \times \text{activity level}$

Once each individual moves to his or her optimal destination the individual “consumes” the meal. The individual banks the caloric content of the meal. The net calorie amount, which is given as calories consumed minus a third of the individuals BMR incrementally updates the person’s weight. BMR is also adjusted to correspond to each iterative change in weight. BMR is calculated using the Harris Benedict Equation (8) and (8’) (Roza and Shizgal 1984). Since the HB equation is also dependent on age, age is also updated every representative year.
2. Results

3.1 Weight Gain Overtime

Figure 3 shows the results of a deterministic model of BMI increase over time holding age constant. Using the male archetype used in the Roza and Shizgal (1984). This model depicts the weight gain of an individual who is initially 1.7m in height, 70kgs and 50 years old. This amounts to an initial BMI of 24.2 and a resting BMR of 1558 kcals/day. Each plot shows weight gain over time from eating the corresponding daily excess calories over 10 years. Unsurprisingly, the individual experiences no weigh gain in the 0+ simulation where the individual eats exactly 1/3 of his initial BMR at each meal. More interestingly are the 200+ and 500+ plots. A daily excess of 200 calories shows quite a significant increase in BMI. In fact over ten years the individual’s BMI increased by 21.29%. This is consistent with empirical findings found by Cutler et al. (2003) who found that a daily excess of just 200 calories leads to significant weight gain. Lastly, the 500+ line shows a very dramatic weight gain. In ten years this person increased
his BMI by 53.22%. This plot is in reference to the common adage that an extra 500 calories a day will lead to a 1 pound increase in weight per week because there are 3500 calories in one pound of fat. However, this graph quite pessimistically suggests otherwise and would propose that weight gain at least initially increases far quicker.

The justification behind this growth trend comes from the fact that BMR is lagged. At \( t \), net calories equals \( kcals_j - BMR_{i,t-1} \). When the individual consumes more than his daily caloric amount he is essentially only equipped with a metabolism, which burns less and as a result gains weight. Over time the difference between BMR and caloric intake decreases as the person’s BMR adjusts to the weight gain. The end result is that \( \frac{\delta BMI}{\delta t} < 0 \). Specifically, we see that BMI grows logistically.

### 3.2 Baseline Model

<table>
<thead>
<tr>
<th></th>
<th>( x_{0,t,s,j} )</th>
<th>( x_{1,t,s,j} )</th>
<th>( \Delta x_{0,t,s,j} )</th>
<th>( \Delta x_{1,t,s,j} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s=1 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td>( s=2 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td>( s=3 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td>( s=4 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td>( s=5 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td>( s=6 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td>( s=7 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td>( s=8 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td>( s=9 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td>( s=10 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td>( s=11 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td>( s=12 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td>( s=13 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td>( s=14 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td>( s=15 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td>( s=16 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td>( s=17 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td>( s=18 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td>( s=19 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td>( s=20 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
<tr>
<td>( s=21 )</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
<td>2.34</td>
</tr>
</tbody>
</table>

Table 4. shows results from the baseline model are consistent with the BMI distribution from NHANES 2009-2010 data after truncating for individuals whose BMI’s were under 13.18.
The results from 1737 Monte Carlo simulations show that the mean and standard deviation of empirical findings are very similar to the average and standard deviation of the NHANES data. And, the statistics for the NHANES data are well within the mean and standard deviation intervals from the 1737 baseline simulations. This provides a sufficient burden of proof that the baseline model is calibrated to NHANES 2009-2010 empirical findings and justifies further exploration of different pricing environments, which may support weight loss.

### 3.2.1 Middle Class Shows Highest Propensity for Obesity in the Baseline Model

**Figure 4.**

\[ BMI_{i,T} \sim income_{i,T} + income_{i,T}^2 + travel\ radius_{i} + \epsilon \]  

Equation (9) is the OLS regression used to calculate the effects of income and mobility on obesity\(^{18}\). The results show that indeed income plays a large role in BMI outcomes.

\(^{18}\) See Appendix for results and summary statistics for income, income\(^2\) and travel radius
are derived from individual level data collected from four baseline simulations and truncated based on the minimum accepted BMI. The constant term was suppressed because of truncation of the data. In the context of this model if income level is zero then consumption is zero and therefore the individual is unable to sustain a body weight and would not be included in this model. Thus, it is reasonable to assume that an income of zero would produce a BMI of also zero. Even the subsidy given to poor households was based on at least some initial minimal income.

Figure 4 depicts the graph of he coefficients on income and income\(^2\) suggest that the relationship with BMI and income is in fact quadratic. Travel radius was also included because access to restaurants (and essentially calories) also impacts weight outcomes. This variable captures neighborhood effects post subsidy. Contrary to my initial hypothesis, middle class individuals exhibit the highest preponderance of obesity\(^{19}\). Essentially, poorer individuals even with inflated purchasing power are still more constrained by their limiting local environments, which are assumed to have inferior access to calories. The income coefficients suggest that poorer individuals are still dominated by an income effect and simply cannot afford enough calories to reach higher obesity status, while richer individuals are wealthy enough to afford to make discretions between meals that are high in satiety versus meals that are high in health. The middle class are then burden with high incidence of obesity because though they can afford and have access to a surplus of calories are still out priced by heath and so are relegated to meals that are predominantly high in satiety, which enables them to gain more weight.

\(^{19}\) The maximum BMI achieved from income alone is 26.84 and corresponds to a income level of $26.58 and is approximately the income level of the 50\(^{th}\) percentile.
3.3 Tax on Carbohydrates

Recall that the original price vector was \( p_{pr} = 3.26, p_c = 0.01, p_f = 0.086 \). In this model I set \( p_c = 0.05 \), effectively placing $0.04 tax on carbohydrates. Table 4 shows that the results are promising as I found that the distribution of mean BMI was significantly less than in the status quo model. In fact the average mean was 26.36, just above the the upper bound for normal BMI. The results are encouraging; however the average mortality rate\(^{20}\) of 5.46% is much higher than what is observed\(^{21}\). Because this phenomenon is not observed in reality it suggests that the model does not fully capturing behavior of the poorest individuals.

3.3.1 Propensity of Obesity shifts to Upper-Middle Class in Tax Model

Using individual level data from 4 simulations I get the coefficients from income and income\(^{2}\) and Table 5. is a graphical depiction. The tax model shows that occurrence of obesity

\(^{20}\)Mortality rate was calculated as the fraction of individuals whose BMI was lower 13.18

\(^{21}\)According to the WHO, the national mortality rate attributed to malnutrition is 0.045% in the United States (http://www.who.int/healthinfo/global_burden_disease/estimates_country/en/index.html)

Saksena 22
shifts to the upper-middle class but at a lower level\textsuperscript{22}. It is calculated that the maximum BMI level of 25.25 corresponds to an income level of precisely $25.12 or roughly the 75\textsuperscript{th} income percentile. Unsurprisingly, the coefficient on travel radius though still positive, decreases in the tax model. This is to be expected as overall availability of calories in the model decreases in response to higher cost of carbohydrates. And, ability to travel to more restaurants becomes less influential in consuming more calories. The shift of burden of obesity to the upper middle class is a result of a more expensive pricing environment. In the baseline model households whose incomes fell between the 50\textsuperscript{th} and 75\textsuperscript{th} percentile were rich enough to buy sufficient amounts of health but lose that flexibility in the taxed environment as their discretionary power is partially reduced.

\textit{4. Limitations of Research}

There are certain limitations of this model, which should be illuminated. First, the ‘mortality’ rate is overestimated when compared to actual deaths attributed to starvation in the United States. Even with the subsidization of the poor, this model does not include any mechanism to simulate food assistance programs, which unlike a subsidy serves as a minimal guarantee for food access to poor individuals. And, perhaps the mortality rate observed in the simulation would be the actual mortality rate in lieu of government assistance.

A logical deduction would be to tax fat since the conversion rate of fat grams to calories is higher and thus less availability of it would induce less weight gain. This is however a naïve construction as consumption of some fats does not necessarily lead to weight gain and even may help to alleviate weight problems. In this model calories from each nutrient are considered equal and the idea of propensity to be converted into and stored as body fat is neglected. For example, glycemic index (GI) is a measure which indicates which foods are more readily stored as fat. For

\textsuperscript{22} See appendix for summary statistics and regression estimates of the tax model
example, table sugar has a GI of 100 and would be considered to be very “fattening” since the
body readily metabolizes and converts it to body fat despite there being no actual consumptive
fat in table sugar itself. Contrastingly, protein has the same gram to calorie conversion as
carbohydrates. However, the way protein is metabolized dictates that only in very larger
amounts of consumption will protein be converted and stored as fat. Unfortunately, GI’s of foods
can only be determined by laboratory methods and so cannot be incorporated in this model.

Lastly, this model also does not address food scarcity. It is more probable that poor
individuals facing high food risk will demonstrate much different behavior then individuals who
expect to eat regularly. As a result poor individuals may exhibit more “feast and famine”
behavior. In other words, since a poorer individual may not expect to be able to eat in regular
increments, in the advent of temporary food access his rational choice maybe to in fact over-
consume above and beyond healthy limits. Since he skips meal events his BMR is likely to also
atrophy making him more susceptible to weight gain. And, in developed countries such as the
United States it is likely that famine intervals are shorter and access to calories during feast
events is much greater. This is not the case in developing countries as net caloric intake maybe
negative because intervals between meals are large enough to compensate for a slowing BMR
and access to calories may also be limited. This may reconcile the high mortality rate in this
model with established\textsuperscript{23} opinions that preponderance of obesity is high among the poor.

5. Conclusion

This paper examines the effects of price environments on incidence of obesity in a
simulation context. The results are promising. The baseline model shows many characteristics
that it is well calibrated to actually BMI data. This serves as a good basis for applying tax
policy. The tax model also showed favorable results as the average BMI of the population

\textsuperscript{23} Though this point has been somewhat refuted as explained earlier.
significantly shifted downward. However, average mortality rates for both the baseline and tax model were higher than what is observed. This discrepancy may be attributable to the fact that the model does not capture the efficacy of social programs that reduce exposure to food insecurity and serve as a minimal guaranteed level of calorie access for the poorest individuals. Secondly, in reality poor individuals may exhibit more feast and famine behavior. Despite the difference in mortality rate, the model shows promise in predicting population obesity outcomes. It is encouraging that such a small tax ($0.04) on carbohydrates shows such dramatic reduction in average obesity. While empirical investigation is stymied by lack of appropriate data (domestically speaking), results from this model show that agent based modeling is a fitting investigative alternative to traditional empirical studies.
Bibliography


Lakdawalla, Darius, and Tomas Philipson. "The Growth of Obesity and Technological Change:


### Appendix

<table>
<thead>
<tr>
<th>Grade</th>
<th>Grade</th>
<th>Unit</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade</td>
<td>Grade</td>
<td>Unit</td>
<td>Unit</td>
</tr>
<tr>
<td>Grade</td>
<td>Grade</td>
<td>Unit</td>
<td>Unit</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>μ</th>
<th>σ</th>
<th>μ</th>
<th>σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>μ</td>
<td>σ</td>
<td>μ</td>
<td>σ</td>
</tr>
<tr>
<td>μ</td>
<td>σ</td>
<td>μ</td>
<td>σ</td>
</tr>
</tbody>
</table>

Saksena 28