Genetic and Economic Interaction in Health Formation: The Case of Obesity.

Pietro Biroli *

September 10, 2015

Latest version here

Abstract

Small genetic differences at birth confer a comparative advantage in health and human capital formation, and can lead to substantial inequality in long term social and economic outcomes. I develop a structural model of health and human capital formation illustrating the dynamic interaction between genetic inheritance and investments in health over the life cycle. Genetic heterogeneity across individuals can change the utility cost of investments and the production function of health, shifting the incentives to invest in healthy habits. Focusing on Body-Mass-Index (BMI) as a measure of poor health, I consider physical activity and food intake as investments in health, and I evaluate their interaction with specific variants in FTO and other genes associated with BMI in Genome-Wide Association Studies. Applying this model to two different datasets, one of British adolescents and one of US adults, I find that Gene-Environment interaction plays a pivotal role in the evolution of BMI. Food intake has a stronger impact on BMI for those individuals with a particular genetic makeup, and yet they tend to display a higher demand for food. The association of variants in the FTO gene with the hypothalamic regulation of food intake gives a biological foundation to the observed differences in healthy investments. This analysis provides an economic framework of health and human capital formation that integrates recent findings in genetics and molecular biology and sheds light on the interdependence between genes and economic choices of investment.

*Department of Economics, University of Zurich, Schönberggasse 1, 8001 Zürich, Switzerland; email, pietro.biroli@econ.uzh.ch; I benefited from helpful comments from Marco Bassetto, Sarah Cattan, George Davey-Smith, Mariacristina De Nardi, Tim Kautz, Gita Khun Jush, Robert Lalonde, Steven Lehrer, Pia Pinger, Maria Rosales-Rueda, Daniel Tannenbaum, Adriaan Ten Kate, Alessandra Voena, as well as from participants at the Annual Health Econometrics Workshop, the IZA/CEPR European Summer Symposium in Labour Economics, the Genoeconomics Workshop in Stockholm, the Lifecycle Working Group at the University of Chicago, the Family Economics Workshop at the University of Chicago, the CPE seminar at the University of Chicago, the IFS lunchtime seminar, and the Chicago Area Workshop in Family Economics. I am particularly grateful to Dan A. Black, Gabriella Conti, Steven N. Durlauf, and James J. Heckman for their continued support.

1 Introduction

Our genetic endowment is determined at conception and stays with us throughout our lives, but every day we choose how to spend time, effort, and money in order to cultivate and shape our natural predisposition. The increase in obesity rates witnessed in the last decades is an illustration of the interaction between genes and lifestyle choices: while the underlying genetic pool remained constant, changes in healthy behaviors have triggered the recent surge in obesity, especially for those individuals who carry a particular genetic variant.[1] The goal of this paper is to understand how choices of investment in health and human capital build on and interact with genetic endowments in order to enable the full flourishing of innate abilities.

To achieve this goal, I introduce a general economic framework that combines recent discoveries of molecular genetics with a dynamic model of investment in health. In this model, the genetic type of an individual can shift health production possibility frontiers as well as her preferences. Genetic variants delineate the set of achievable combinations of choices (healthy behaviors) and outcomes (health), and confer to certain individuals a comparative advantage in the production of health. This framework sheds light on how genetic endowments affect decisions about investments in health and human capital. Focusing on Body Mass Index (BMI) as a measure of obesity, I show how genes interact with the environment in two ways: they can shift the implicit costs of investments in diet and exercise; and they can change the productivity of these inputs in the formation of BMI. These genetic effects induce a change in the optimal level of investments, and in the long-run level of BMI. The model is general enough to be applied to different types of human capital, investments, and genes.[2] For example, leveraging the recent results of Rietveld et al. (2014b)[3] the analysis can be extended to model how cognition

[1] See Rosenquist et al. (2014)

[2] I follow the suggestions of Moffitt et al. (2005) and Purcell (2002) to select precise measures of outcomes, genes and environments. They point out that gene-environment interactions (GxE) can be detected using twin models which specify genes and environments as latent variables; however this strategy suffers from low power, is sensitive to non-normality, and does not shed light on the underlying processes. Using well defined measures of genes and environments is more sensible both from a statistical perspective - it provides the most power for detecting GxE - and an analytical perspective - it sheds light on the causal links connecting endowment, choices, and the final outcome

[3] These results build on previous work on cognition and education by Rietveld et al. (2013, 2014a), and have been replicated in a cohort of children by Ward et al. (2014)

and investment in education can be influenced by one's genetic endowment.

Obesity has become of prime importance due to the dramatic rise in body-mass witnessed in the last decades, especially in children. Ogden et al. (2002) and Ogden et al. (2012) show how obesity rates of US children aged 2 to 5 doubled from 1970 to 2000, going from a prevalence of 5% to 10.4%, and plateaued at 12.1% in 2010; obesity rates for US children aged 6 to 19 tripled in the same time frame, going from 5% in 1970 to 15% at the turn of the century, reaching 18.2% in 2010.[4] A similar trend occurred in England.[5] Obesity is the second most important cause of premature death in the US,[6] and its negative health consequences are far-reaching, from type 2 diabetes to coronary heart disease and certain types of cancer.[7] Furthermore, Cawley (2010) shows how both direct and indirect costs of obesity are considerable: the medical cost of childhood obesity in the US is $14.3 billion a year, and the figure for adults is 10 times greater, with an estimated $147 billion spent in obesity related illnesses.[8] The indirect costs of obesity range from delayed skills acquisition, to lower wages, job absenteeism, and lower productivity.

Obesity is well suited for this analysis because it has a strong genetic basis, yet it can be influenced by individual choices of diet and physical exercise.[9] I use a variant of

[4]There is some evidence that obesity rates has stopped increasing in the last few years, especially for preschool children, see Ogden et al. (2014) and Wabitsch et al. (2014). More recent statistics can be found at `http://www.cdc.gov/obesity/data/childhood.html`

[5]Stamatakis et al. (2010) look at difference between 1995 and 2007, showing how obesity rates increased from 2.9% to 6.4% for children aged 2 to 5, more than doubled from 3.3% to 7.3% for children 6 to 10-years-old, and rose from 2.7% to 4.8% for adolescents aged 11 to 18.

[6]Rashad (2006)

[7]Daniels et al. (2009) connects childhood obesity to the risk of developing metabolic syndrome, hypertension, dyslipidaemia, non-alcoholic steatohepatitis, and obstructive sleep apnea. Singh et al. (2008) tracks childhood obesity into adulthood. For the negative consequence of adult obesity, see LeBlanc et al. (2011), Whitlock et al. (2009) and the references therein.

[8]Finkelstein et al. (2009) obtain these estimates using 2008 data from the National Health Expenditure Accounts and comparing the overall medical cost of obese and non-obese people. A similar approach using data from the Medical Expenditure Panel Survey (MEPS), which does not include medical spending for people residing in institutions, leads to an estimate of $85.7 billion. Cawley and Meyerhoefer (2012) use the Medical Expenditure Panel Survey with an instrumental variable approach and obtain an even higher figure of $210 billion. Finally, using a meta-analysis Finkelstein et al. (2014) recommend an estimate of $19,000 as the incremental lifetime medical cost of obesity.

[9]Surveying the literature, Yang et al. (2007) report that 16% to 85% of Body Mass Index is 'heritable' and related to genetic similarities among twins. Sandholt et al. (2012); Speliotes et al. (2010) identify a list of genetic variants that have been associated with obesity. In a historical perspective of genetic-obesity research, Jou (2014) reviews many studies of twins, adoptees, and experimental overfeeding, concluding that "Genetic predispositions, in tandem with the development of food envi-

the FTO gene in this paper because of the importance of FTO in determining obesity, the wide prevalence of the risky allele, and the known established biological connections between FTO and appetite. The minor allele of the rs9939609 single nucleotide polymorphism (SNP) in FTO gene was first related to susceptibility for obesity in genome-wide association studies (GWAS) by Frayling et al. (2007). This genetic variant was found to be associated with a 20%-30% increase in the risk of obesity and a 1-1.5 kg increase in body weight.[10] The explanatory power of a single genetic variant is expected to be modest, especially when dealing with a complex disease such as obesity which is determined by multiple genes and various lifestyle choices. Many other genes have been associated to obesity using GWAS, but "FTO remains the gene with the most robust association and greatest effect size." (Yeo and O'Rahilly (2012)). Furthermore FTO is quite common, with a minor allele frequency close to 50% in genetically diverse populations. [11] Using the words of Tung and Yeo (2011), "FTO could possibly be influencing the BMI of up to half the world's population!" Finally, established findings in molecular genetics link FTO to hypothalamic regulation of appetite and food intake,[12] suggesting that this genetic type increases the cost of following a strict diet without altering the incentives to engage in physical activity.

I use a novel epidemiological dataset that combines DNA-assays with precise information on children Body Mass Index, their dietary choices, and their level of physical activity. I find that the predictions of the model are borne out by the data: the genetic endowment of the child changes the structural parameters of the production function of obesity, as well as the implicit cost of investing in diet. Specifically, I find evidence of a productivity effect of the FTO genetic variant, which interacts with the level of caloric intake and increases the risk of obesity especially for abundant eaters. In other words, there is evidence of gene-diet interaction. Even when facing such greater risk,

ronments that facilitate overeating and built environments requiring minimal energy expenditure, may help explain why so many Americans are obese today"

[10]This findings were then replicated by multiple authors using different datasets, see Dina et al. (2007); Timpson et al. (2008)

[11]Using HapMap population, Kilpeläinen et al. (2011) estimates that 74% of individuals of European descent, 76% of individuals of African-American descent, and 28%-44% of individuals of Asian descent carry one or more copies of the FTO risk allele.

[12]See for instance Speakman et al. (2008) and Wardle et al. (2008). A more detailed description of the findings is in section (2).

children endowed with at least one A-allele in the rs9939609 FTO genetic variant still end up consuming more calories. This is consistent with a genetic effect on the cost of investment in diet. There is no evidence of a significant interaction between this genotype and physical activity. To summarize, the FTO genetic variant is associated with being at greater risk of obesity for those who ingest a lot of calories and at the same time with consuming more food. These two effects jointly explain why individuals with this particular genotype tend to be more obese. However this result is conditional on investment choices in diet and exercise. Higher levels of investment in exercise and greater effort in maintaining a lower caloric intake can offset the negative consequences of being born with a particular genotype.

These results are robust to different specifications of the production function of obesity. Using different measurements of obesity or investments does not change the qualitative findings. Since obesity is a complex disease that is influenced by multiple genes, a polygenic approach is considered as a robustness check. I construct a composite measure of genetic predisposition to obesity, using information from 24 genes that have been consistently related to body fat, and confirm the main results of the analysis. Furthermore, these results are replicated using two separate datasets, one of UK children born in the 1990s and the other of US adults born over the last century.

My contribution lies at the crossroad of three different strands of literature. First, I take part in the debate on genes vs. environment by introducing an economic framework to understand how individual choices are shaped by genetic variants and economic environments. This long-standing debate was first introduced by Mulcaster (1582) and Galton (1874) and recently discussed by Lundborg and Stenberg (2010); Heckman (2007); Rutter (2006). Secondly, my work is embedded into the novel research on Genoeconomics spearheaded by Benjamin et al. (2007), Beauchamp et al. (2011) and Conley et al. (2014), who investigate the genetic determinants of economic outcomes.[13] I also relate to the studies on Mendelian randomization initiated by Davey Smith (2003) and Davey Smith and Ebrahim (2004), who use genetic variants as instrumental variables.[14]

[13] Benjamin et al. (2014) also analyze the way genetic variants enter a structural economic model of healthy habits, but they focus their attention on nicotine-receptor genes and smoking.

[14] These methods have been extensively used and discussed, see for instance Norton and Han (2008);

Finally, by using genetic variants as a biological foundation for individual heterogeneity, I contribute to the literature on the dynamics of health and obesity, initiated by Grossman (1972) and extended by Scholz and Seshadri (2013); Cawley (2010); Cutler et al. (2003); Lakdawalla et al. (2005).

The analysis suggests that, although many genetic loci have been associated with higher levels of BMI, obesity rates are strongly determined by the interaction between genes and environment, and behavioral and economic choices can prevent and curtail the insurgence of adiposity. Although genes are immutable and are strongly connected to differences in obesity, policies targeted at children that promote healthy behaviors, such as diet and regular physical activity, could be very effective in reversing the recent trend in obesity rates. Furthermore, carefully planned policies can leverage the interaction between genes and environment in order to reduce the lifetime inequality determined by genetic differences at birth.

The rest of the paper is organized as follows. Section (2) describes in the detail the model. Section (3) introduces the data and the empirical estimation. Finally, section (4) draws some conclusions and discusses about future avenues of research.

2 The Model

Following the seminal work of Grossman (1972) and Grossman (2000),[15] I develop a structural model of health and human capital formation that takes into account the dynamic interaction between genetic inheritance and healthy choices of investment over the life-cycle. This model can be easily generalized to various definitions of human capital H, investments I_k, and genetic markers g. However, precise measurements for each of these three fundamental components must be specified in order to test the theory and its implications. The choice of these three components must be thoughtful

von Hinke Kessler Scholder et al. (2013, 2011). Fletcher and Lehrer (2011) discuss the importance of looking at within-family variations. For a discussion of potential violations of the I.V. assumptions, see Cawley et al. (2011) and Fang (2013).

[15] See also the work of Lakdawalla et al. (2005); Lakdawalla and Philipson (2009); Dustmann and Windmeijer (2000); Galama and Van Kippersluis (2010); Galama (2015); Cunha and Heckman (2008); Conti and Heckman (2010).

and based on the existing evidence that connects investments and genotype to the final outcome of interest.[16] I use body mass as measure of H, food consumption F and exercise E as proxies for I_k, and FTO as assays of g.

Utility $U(B_t, F_t, \ell_t, c_t; g)$ is derived from non-food consumption c_t, food consumption F_t, leisure ℓ_t, and health as measured by Body-Mass-Index B_t, and it is separable in consumption and monotonically increasing in the first three arguments.[17] The future stock of health $B_{t+1} = I(F_t, E_t; g) + (1 - \delta_t)B_t + \varepsilon_t$ depends on the current stock B_t, which depreciates over time (δ_t), is subject to a health shock ε_t, and can be increased by proper investments I_t.[18] Investments in health are achieved through choices of dieting and physical activity, so that the future stock of health B_{t+1} is obtained by combining calories F_t and exercise E_t, $I_t = I(F_t, E_t; g)$, where the function $I_t(.)$ is continuous, monotone, and concave in its arguments.[19] In each time period, the current level of the health stock determines the amount of productive time Ω_t, which can be allocated to either leisure ℓ_t or investing in exercise E_t; Ω_t is the service flow per unit of health stock. Each period, agents are endowed with income Y_t that is devoted to buying general consumption c_t and food consumption F_t. The latter can can be considered a composite of two types, nutritious food NF_t with a lower caloric content, and non-nutritious food UF_t with a higher caloric content but lower relative cost. The overall expenditure on food will be $p_{F_t}F(UF, NF) = p_{UF_t}UF_t + p_{NF_t}NF_t$, so that $Y_t = p_f F(UF_t, NF_t) + c_t$.[20]

[16] See Moffitt et al. (2005) for suggestions on how to approach the investigation.

[17] Conceptually, utility is also increasing in health, which is a non-linear function of body mass. According to WHO standards, health is higher when adult BMI is between 18.5 and 25. Therefore low utility is derived from poor health due to either very high or very low weight. Additionally, the model can accommodate self-perception and "ideal-weight" criteria.

[18] Following Ehrlich and Chuma (1990) and Galama et al. (2012b) I do not assume constant returns to scale as in Grossman. I also follow their assumption that only the past stock of health is relevant for current health. In the case of obesity, this rules out Jules Hirsch's hypothesis that ever being obese can change your metabolic rates, requiring more willpower to maintain a reduced weight, see Hirsch (2003). Therefore the depreciation rate δ can be thought of as a metabolic rate, or the energy expenditure needed to carry over to the next period; for example in the case of a 150 pound man, the basal metabolic rate is about 1500 calories.

[19] These choices are bounded below by zero and above by some physical limit; however I do not consider the extreme cases of starvation or bedridden individuals, and therefore such limits are never binding.

[20] This simple model abstracts from saving and capital accumulation choices, so that income spent on food is exogenously given each period. Y_t can vary over time, but does not react to body-mass-index or food choices. The underlying assumption is that saving decisions and food consumption are not strongly influenced by each other; this assumption seems reasonable since I am considering wealthy

The total amount of food intake is $F_t = \kappa UF_t + NF_t$, where κ is the caloric content of nutritious relative to non-nutritious food. The fact that non-nutritious food has a higher caloric content is captured by $\kappa > 1$.

The state variables characterizing each individual in any given time period are income Y_t, the health stock B_t, the health shock ε_t, and the genetic endowment g. The choice variables are non-food c_t and food consumption F_t, as well as exercise E_t.[21] The agent's problem can be written as follows:

$$V(B_0, Y_0, \varepsilon_0; g) = \max_{NF_t, E_t} \sum_{t=0}^{T-1} \beta^t EU\left(B_t, F_t, \ell_t, c_t; g\right) \tag{1}$$

$$s.t$$

$$\Omega(B_t) = \ell_t + E_t$$

$$Y_t = p_{F_t} F(UF_t, NF_t) + c_t$$

$$B_{t+1} = I(F_t, E_t; g) + (1 - \delta_t)B_t + \varepsilon_t$$

The innovation of the model is that genetic variants can affect the accumulation of health and human capital, and they do so in two ways: changing productivity and changing preferences. They can influence the production function of health $I(.; g)$: genetic variants can change how productively inputs are converted into outputs can be called the genetic *productivity effect*. This effect is related to the literature on Gene-Environment interaction (GxE).[22] For example, consider a genetic variant that increases

countries where food intake is always attained and other types of consumption drive asset accumulation. Adding a state variable tracking the evolution of assets is straightforward in the theoretical model, but complex for the empirical counterpart given the absence of good asset or saving variables in the datasets used. Similarly to the case of *two-stage budgeting*, in the case of adult individuals Y_t can be considered as the exogenous share of total income that is mentally devoted to food consumption; in the case of children and adolescents, Y_t is the amount of resources that the family devotes to the child's nutrition. In both cases, the choice made by the agents rests on the *composition* of expenditure between nutritious and non-nutritious food.

[21] Non-nutritious food consumption can be easily backed out from the budget constraint, so that $UF_t = (Y_t - NF_t)/p_t$

[22] For the same level of inputs (environment), a different level of output (phenotype) is obtained depending on the genotype of the individual. To use the words of Plomin et al. (1977): "Genotype-environment interaction refers to the possibility that individuals of different genotypes may respond differently to environments." In the words of the model, this means that the effect of the investments I_k depends on the gene, or that $\frac{\partial B}{\partial I_k}$ is a function of g.

the process of fat-storage. For the same amount of food intake, this genetic variant will change the observed level of body mass.[23] Another way of looking at this same effect is to consider the differences in investments (food intake) needed to attain the same final outcome (BMI).[24]

A second channel is the effect of genetic variants on preferences $\partial U(.;g)/\partial F_t$. Achieving a certain level of investment could be easy and effortless for a child with a particular genotype, but very hard and costly for somebody with a different genetic endowment. In the model, this is captured by the parameters governing the utility function. Since this channel influences the implicit cost of investment, it can be labeled the genetic psychic *cost effect*.[25] This effect can be loosely related to Gene-Environment correlation (rGE).[26] For example, consider a genetic variant that increases the preference for food intake: this will make following a diet more strenuous, effectively increasing the cost of such investment, and inducing higher body mass.[27]

With the respect to the information set available to the agents at the time of the decision, I assume that the agents know the relevant parameters of the production

[23]For a more general example, consider a model dealing with human capital in the form of cognitive ability and investments in terms of schooling. A genetic variant that facilitates the creation of neural connections has a *productivity effect*: for the same amount of investment in schooling, this genenetic variant will increase the cognitive ability of the child. A more famous example comes from the research of Caspi et al. (2002), who show how variations in the MAOA gene moderate the effect of childhood maltreatment on adult anti-social behavior. In this case maltreatment I_m is a (negative) input in the production function of antisocial behavior H_b, a (negative) measure of human capital. We have that I_m is very effective in producing H_b only in presence of high MAOA activity.

[24]For example Gobet and Campitelli (2007) focus on chess players and investigate the different amount of practice hours needed to achieve the level of "master."

[25]More appropriately, this channels constitutes a genetic *utility cost effect* associated with changes in investment choices. The implicit cost of investment will contain a term $\partial U/\partial F_t$ that depends on genes.

[26]Plomin et al. (1977) say that Genotype-Environment correlation "occurs if different genotypes are selectively exposed to different environments." In the words of the model, this means that certain investments I_k are more prevalent for children with a particular genotype. This is a statement about equilibrium levels of investment, which can have many causes. However, if the subjective cost of investment is lower for certain genotypes, we would expect them to be "exposed" to higher levels of I_k.

[27]A more famous example comes from the so-called 'Asian glow' syndrome. Davey Smith (2010) shows how a null variant in the *ALDH2* genotype, very common in men of Asian descent, renders the consumption of alcohol unpleasant because of facial flushing, palpitations, and drowsiness. In terms of a more general model of health, this genetic variant is related to an increase in the cost of drinking, a negative investment in one's health. Indeed, this genotype is related to lower consumption of alcoholic beverages, and consequently to a lower risk of liver cirrhosis, HDL cholesterol levels, and blood pressure.

function and observe the cost of the investments. other words, the agents know how costly it is to achieve a certain level of investment (knows the prices and $I(.)$), and how useful such investments are in the formation of human capital; however, they do not need to know the particular alleles of the genetic code.

Analogous to the household production model of Becker, the genetic endowment would be considered as an 'environmental variable' which influences the household production function. [28] The genotype of an individual changes the parameters of the model, shaping the production possibility frontier and the incentives to investment faced by the agent.

Indeed, notice how both genetic effects are associated with the welfare of individuals. The cost effect reduces the cost of investments. The productivity effect increases the human capital produced with the same level of inputs. Both effects are related to an improvement in the production possibility frontier, so that more human capital can be attained spending the same amount of resources, or equivalently the same level of human capital is achieved spending less time and money investing.

In the following sections, first I will consider a simple three period model to fix ideas and better understand the mechanisms at work; then I will discuss in detail the effect of the genotype g on the incentives to invest and on the accumulation of health. The description of the solution of the dynamic model using the Bellman equation can be found in the appendix B.

2.1 Three period model

To fix ideas and understand the main mechanisms at work, consider a three-period model of investment in health ($T = 3$). The agent is assumed to maximize

$$V(B_1, Y_{\{t\}}; g) = \max_{NF_t, E_t} EU(B_1, F_1, \ell_1, c_1; g) + \beta EU(B_2, F_2, \ell_2, c_2; g) + \beta^2 EU(B_3, F_3, \ell_3, c_3; g)$$

$$s.t. \Omega(B_t) = \ell_t + E_t$$

$$Y_t = p_{F_t} F_t + c_t$$

[28] As explained in (Becker, 2007, p.48), such variables "reduce the cost of producing commodities, and thus would expand opportunities, even if the full income were not affected."

$$B_{t+1} = I(F_t, E_t; g) + (1 - \delta_t)B_t$$

where B_1 and the sequence of incomes $Y_{\{t\}}$ are given.

In the final period, there is no incentive to invest in future health, so that exercise will be zero, $E_3 = 0$, and expenditure will solve $\partial U_F(F_3) = p_{F_3}\partial U_c(Y_3 - p_{F_3}F_3)$.[29] The optimal food consumption in period 2 is given by:

$$\begin{aligned} U'_{F_2}(.;g) &= p_{F_2}U'_{c_2}(.;g) + \beta E\left[-U'_{B_3}(.;g) - U'_{\ell_3}(.;g)\frac{\partial \Omega}{\partial B_3}\right] I'_{F_2}(.;g) \\ &= p_{F_2}U'_{c_2}(.;g) + \beta E\left[\varphi_{B_3}\right] I'_{F_2}(.;g) \end{aligned}$$

The left-hand-side of the equation captures the genetic cost effect $U'_F(B_2, F_2, \ell_2) = \partial U/\partial F_2$, while the term $\partial I/\partial F$ in the right-hand-side capture the genetic productivity effect. The split between nutritious and non-nutritious food is determined by income and relative prices, such that $NF_t = F_t - \kappa UF_t$ and $UF_t = \frac{Y_t - c_t - P_{NF}F_t}{p_{UF} - \kappa P_{NF}}$. Optimal investment in exercise must satisfy:

$$-U'_\ell(B_2, F_2, \ell_2) = \beta E\left[\varphi_{B_3}\right] \frac{\partial I(F_2, E_2; g)}{\partial E_2}$$

The optimal choice of food consumption in period 1 must satisfy the following:

$$U'_F(B_1, F_1, \ell_1) = p_{F_1}U'_{c_1} + \beta E\left[\varphi_{B_2} + \beta(1 - \delta_2)\varphi_{B_3}\right] \frac{\partial I(F_1, E_1; g)}{\partial F_1}$$

where this time the expected marginal loss of health takes into account also the spillover into future health, φ_{B_3},discounted by the depreciation rate of health and the discount rate $\beta(1 - \delta_2)$. This is due to the fact that an increase in the health stock

[29] Since non-nutritious food is a more efficient way of obtaining calories in period 3 as long as $p_t < \kappa$, we have that $NF_3 = 0$ and $UF_3 = (Y_3 - c_3)/p_{F_3}$. The underlying assumption is that nutritious food is more expensive than non-nutritious one, in terms of caloric content, so that $p_t/\kappa < 1$. If that was not the case, then every time period the budget would be spent on nutritious food, which both has positive health benefits and also would satisfy more effectively the agent's preferences. Another option would be to have both types of food to enter the utility function separately. In this case the FOC would require the ratio of marginal utilities from the two consumptions to equal the relative price, $p_3 = \frac{\partial U/\partial UF_3}{\partial U/\partial NF_3}$

tomorrow, B_2, has an impact on the health stock in the future, B_3, unless there is full depreciation ($\delta_2 = 1$). For any additional year ahead of the current period, there is a marginal increase in the benefit of investing in the stock of health capital.

Comparing a change in the cost of food consumption in period one or in period two, we can see how a decrease in the implicit price of food today has a greater effect on investments and health. This is due to the fact that a higher level of investment today has an impact on both health today and health tomorrow, because of the law of motion of health capital. Unless the health stock fully depreciates in one period, by investing today we will rip the benefits in all following periods.

2.2 The Effect of Genes

Consider how the genotype of the agent influences the decision making process. In terms of their impacts on health, in the model genes do not change the marginal benefit of health, but they do have an impact on the marginal cost of health. Such impact is due to both the fact that genes can change the productivity of investments, shifting $\frac{\partial I}{\partial F_t}$, and that genes can influence the cost of achieving such investment, shifting $\frac{\partial U}{\partial F_t}$.

Consider FTO: various studies in molecular and human genetics have shown how FTO is associated with obesity through the regulation of appetite and hunger (energy intake). It does not seem to influence physical activity or calories burnt (energy expenditure).[30] For example Fredriksson et al. (2008) use mice-models to analyze the biological functioning of FTO: they find that this gene is highly active in certain parts of the brain that regulate feeding impulses and appetite, notably in the hypothalamus. This was particularly true in the brain of mice who had been starved for two days.[31]

[30] Tung and Yeo (2011), Yeo and O'Rahilly (2012), and Fawcett and Barroso (2010) overview the evolving importance of FTO in the field of the genetics of obesity, and the various discoveries of its biological functions. For more detailed analysis, see Speakman et al. (2008); Fredriksson et al. (2008); Tung et al. (2010) who use animal models; Wardle et al. (2008); Timpson et al. (2008); Cecil et al. (2008) show results in different human populations.

[31] They find that "detailed in situ hybridization analysis in the mouse brain showed abundant expression in feeding-related nuclei of the brainstem and hypothalamus, such as the nucleus of the solitary tract, area postrema, and arcuate, paraventricular, and supraoptic nuclei as well as in the bed nucleus of the stria terminalis. [...] The FTO was significantly up-regulated (41%) in the hypothalamus of rats after 48-h food deprivation." They conclude that "These results are consistent with the hypothesis that FTO could participate in the central control of energy homeostasis."

Similar results are reported by Olszewski et al. (2009) and Tung et al. (2010), finding significant changes in the activity level of FTO in the hypothalamus of rats and mice experimentally deprived of food.[32] Turning to evidence from human studies, Cecil et al. (2008) analyze 2,726 Scottish children, 4 to 10 years of age, and find that the "A allele [of the rs9939609 FTO genetic variant] was associated with increased energy intake independently of body weight"; however it had no visible effect on their resting energy expenditure and basal metabolic rate. Analyzing the same dataset used in this paper, Timpson et al. (2008) find a strong effect of the A-allele on increased total food intake and total fat intake of children with similar body mass.

In terms of the model this means that being a carrier of the FTO A-allele changes the utility derived from food consumption: $\partial U/\partial F(A_{FTO}) \gtrapprox \partial U/\partial F(T_{FTO})$. This will increase the marginal cost if investing in diet: there is a genetic cost effect.[33] The increase in the cost of one investment will induce a change in the optimal allocation of both food consumption F_t and time spent exercising E_t. Assuming no effect of FTO on the productivity of investments, the model predicts that being born with at least one A allele in the rs9939609 gene-polymorphism leads to a lower level of investment in diet and, consequently, to a lower level of health (higher BMI). The consequences for the level of exercise are not straightforward but depend on the substitution between the two investments.

Although the case of FTO has been analysed in full detail, very similar physiological functions have been found for other obesity-related-genes, such as MC4R, BDNF, SH2B1.[34]

[32]The level of messenger RNA (mRNA) and expression of a gene is related to its biological activity. Olszewski et al. (2009) report that "FTO mRNA is present mainly in sites related to hunger/satiation control; changes in hypothalamic FTO expression are associated with cues related to energy intake rather than feeding reward. In line with that, neurons involved in feeding termination express FTO." Tung et al. (2010) also experimentally manipulate the expression level of FTO in the Arcuate Nucleus of the hypothalamus and find that a 2.5-fold overexpression induces a 14% reduction in average daily food intake, while knocking down FTO expression by 40% increases food intake by 16%. They conclude that "The regional specific manipulation of FTO expression provides further support [...] that FTO itself can influence energy homeostasis by having direct effect on food intake.

[33]The biological evidence in this regard is strong, but only suggestive. It is important to be cautious in its interpretation and not jump to conclusions. For this reason I use the approximation$\gtrapprox$. Tung and Yeo (2011) say: "what is the physiological function of FTO and what is its role in the control of energy balance? In short, we still do not know for sure."

[34]See Huszar et al. (1997); Govaerts et al. (2005); Qi et al. (2008); Valette et al. (2012) for mice-

To conclude, the wealth of evidence discovered by genetics and molecular biology can help shed light on the expected sign and magnitude of the parameters of the economic model, and guide the empirical exercise.

3 Empirical Results

3.1 The Data

To bring the model to the data, I use both a prospective cohort of children in the UK, and a sample of American adults. The main dataset is the Avon Longitudinal Study of Parents and Children (ALSPAC), an ongoing investigation on the health and development of young children (www.alspac.bris.ac.uk). An extremely rich dataset collected by epidemiologic researchers from the University of Bristol, the ALSPAC follows prospectively a cohort of pregnant women living in a district in the former county of Avon, in the South West of England, with an expected delivery date between April 1991 and December 1992.[35] 14,541 women, contributing 14,062 live births, were initially recruited, and genetic information is available for 8,317 children.

The qualitative findings from this first analysis are then replicated using the Offspring Cohort of the Framingham Heart Study (FHS) (www.framinghamheartstudy.org/).[36] One of the longest running, multigenerational longitudinal study in medical history, in 1971 the FHS started collecting information on the children of the original cohort, examining over 8 waves of clinic visit more than 5 000 individuals born over a period of 60 years (from 1905 to 1965).[37]

Health and lifestyle data were collected through regular questionnaires, as well as

knock-out models as well as human evidence of the relation between the melanocortin-4 receptor (MC4R) and excessive feeding (hyperphagia), high levels of insulin and blood sugar (hyperinsulinemia and hyperglycemia), and increase in food consumption; Gray et al. (2006); Unger et al. (2007) highlight the links between inhibition of food intake, energy homeostasis and the expression of brain-derived neurotrophic factor (BDNF) in the hypothalamus; Bochukova et al. (2010); Li et al. (2007); Ren et al. (2007) explain the relation between leptin, the SH2B1 gene, and eating and obesity. Finally Beckers et al. (2009) overviews the literature on the genetic basis of the leptin-melanocortin pathway to obesity.

[35] See Fraser et al. (2013); Boyd et al. (2013)

[36] Replication of results is quite standard in the genetic literature. Beauchamp et al. (2011) discuss the importance of replication for social-scientist using genetic information.

[37] See the description of the data provided in Govindaraju et al. (2008).

medical and educational records. Anthropometric, physical activity, and food intake measures were obtained during research clinic visits for both datasets.[38]

The summary statistics of the main variables used are found in tables (1-2). As expected, the A-allele of the FTO gene is related to higher levels of BMI for both males and females, and for higher levels of food intake for males. Very similar results can be found when looking at different measures of adiposity, physical activity, and food intake.[39]

Table 1: Summary Statistics by age, gender, and genotype

	Body Mass Index				Sedentary Hours			
	Female		Male		Female		Male	
Age	T-allele	A-Risky	T-allele	A-Risky	T-allele	A-Risky	T-allele	A-Risky
8	16.25	16.42	16.06	16.13	.	.	.	.
	(4.71)	(4.57)	(3.37)	(3.59)	.	.	.	.
	[0.07]	[0.05]	[0.06]	[0.04]	.	.	.	.
11	18.50	18.99	18.17	18.62	7.18	7.25	6.89	6.98
	(10.39)	(10.80)	(8.56)	(10.29)	(1.19)	(1.21)	(1.27)	(1.45)
	[0.10]	[0.08]	[0.09]	[0.07]	[0.04]	[0.03]	[0.04]	[0.03]
13	20.41	20.87	19.74	20.08	8.26	8.24	7.73	7.77
	(11.84)	(12.56)	(10.29)	(11.68)	(1.32)	(1.31)	(1.50)	(1.54)
	[0.12]	[0.09]	[0.11]	[0.09]	[0.05]	[0.03]	[0.05]	[0.04]

Mean of Body Mass Index (BMI kg/m^2), sedentary hours, and Kilocalories (in thousands), by age, gender, and FTO genotype. Sample variance in parenthesis; mean standard-error in brackets.

3.2 Evidence of interaction between genes and investment

Before estimating the structural model in section (3.4), I examine the interaction between genes and investment in the raw data. In section (3.2.1) I describe and analyze the raw data in more detail; in section (3.2.2) I estimate a reduced form version of the law of motion of health, looking for evidence of a productivity effect of genes in the formation of health; then in section (3.2.3) I consider how genetic variants can impact the chosen level of investments, looking for evidence consistent with a genetic cost effect.

[38] Additional information on the particular measured used can be found in the appendix (C).
[39] See table (11) in the appendix.

Table 2: Summary Statistics by age, gender, and genotype

	Kilocalories				Whole Sample		
	Female		Male				
Age	T-allele	A-Risky	T-allele	A-Risky	BMI	Sed	Kcal
8	1.64	1.64	1.75	1.79	16.23	.	1.71
	(0.08)	(0.08)	(0.09)	(0.11)	(4.08)	.	(0.10)
	[0.01]	[0.01]	[0.01]	[0.01]	[0.03]	.	[0.00]
11	1.75	1.78	1.92	1.97	18.64	7.10	1.86
	(0.13)	(0.12)	(0.15)	(0.16)	(10.23)	(1.31)	(0.15)
	[0.01]	[0.01]	[0.01]	[0.01]	[0.04]	[0.02]	[0.01]
13	1.77	1.76	2.12	2.15	20.34	8.02	1.95
	(0.21)	(0.18)	(0.30)	(0.27)	(11.92)	(1.47)	(0.27)
	[0.02]	[0.01]	[0.02]	[0.01]	[0.05]	[0.02]	[0.01]

Mean of Body Mass Index (BMI kg/m^2), sedentary hours, and Kilocalories (in thousands), by age, gender, and FTO genotype. Sample Variance in parenthesis; mean standard-error in brackets.

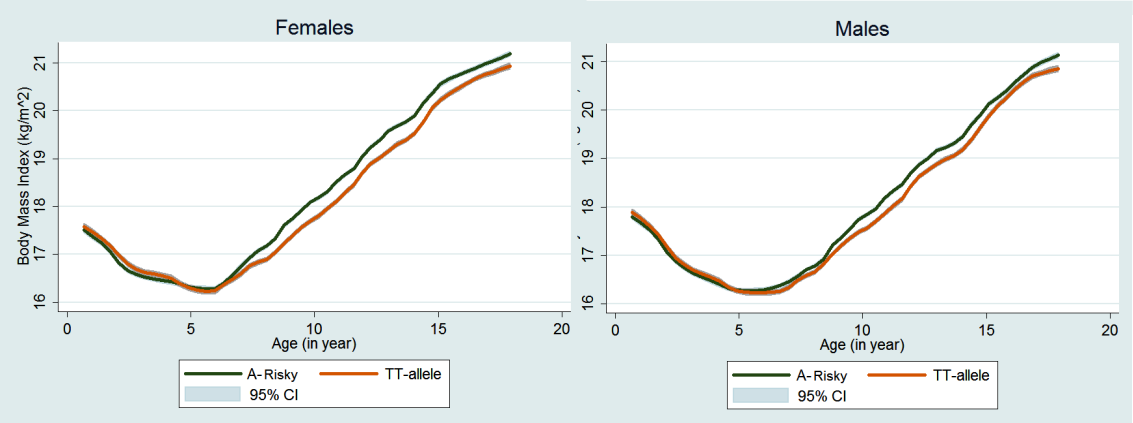
3.2.1 The Raw Data

I split the sample in two: those children who carry at least one A-allele in the FTO gene (homozygous AA carries, or heterozygous AT), who represent 63.2% of the sample; and those who don't (homozygous TT carriers).[40] Figure 1 depicts the evolution of body mass index from birth to age 18; in the first 5-6 years of life there is no statistical difference in obesity between the two types, while the distance between the two groups increases as children get older. The FTO-genotype has a significant effect on Body Mass Index, but such effect is not present since birth: it arises only as the child grows. A very similar result is reported by Rzehak et al. (2010), who find no difference in BMI up to the first three years of life, and then a significantly higher BMI for the carriers of the A-allele. Such a widening gap is consistent with the hypothesis that FTO itself is not sufficient to induce obesity, but rather that the impact of the genetic variant becomes pronounced as the effect of environment accumulates over time.

The genetic productivity effect predicts a different level of output B_t using the

[40]By splitting the sample in two I am making the assumption of dominant genetic effects for the risky A-allele, instead of considering an additive genetic model with three distinct genotypes AA, AT, TT. This assumption is warranted by the small fraction of individuals homozygous for the short AA allele, and the fact that running the model separately for the three genotypes I did not find significant differences between the AA and AT populations.

Figure 1: Evolution of Body Mass Index



Nonparametric local-mean smoothing using Epanechnikov kernel and Silverman's Rule-of-Thumb bandwidth. Combining information from successive clinical visits, age 0 to 18; excluding outliers in the top and bottom 5% of the BMI distribution.

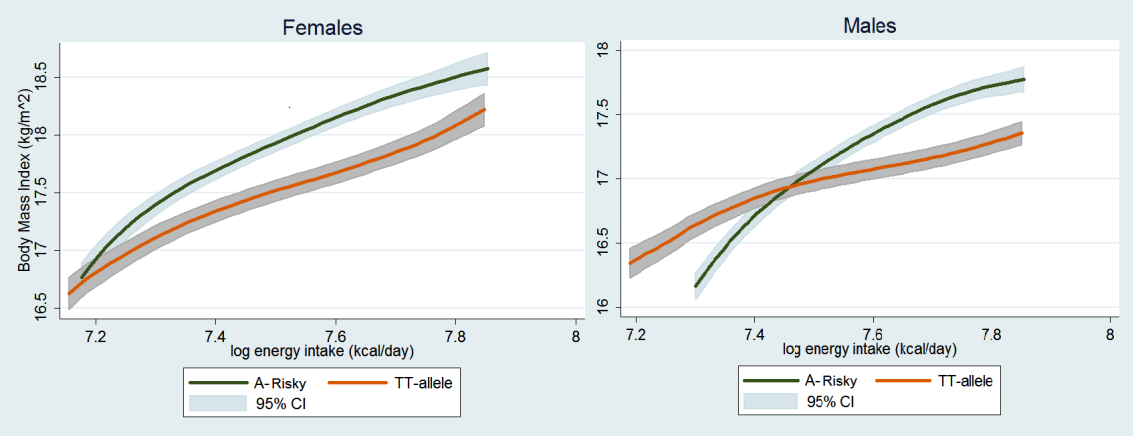
same levels of inputs $I_t = (F_t, E_t)$, depending on the genotype of the individual. At first consider again the raw data, simply plotting the average BMI of the children for different levels of investments. This can give an idea of the shape of the production function $f(.)$.

Analyzing food intake F_t by gender and genetic endowment, I investigate the relationship between investment in diet and the obesity of children aged 10 to 14 years old.[41] Figure 2 shows the relation between the logarithm of the total amount of food intake (kilo-calories per day) and the BMI using a local smoothed average. Not surprisingly, a higher food intake is related to higher levels of BMI. The most interesting feature is the significant difference in the slopes depending on the genotype of the child. It is apparent that $f'_{F_t}(g = A) \geq f'_{F_t}(g = T)$. Furthermore, the two slopes intercept at low levels of caloric intake: genetic differences between children lead to differences in BMI only when they are abundant eaters. The impact of genes is conditional on a particular environment. The effect on obesity is evident only when both genetic variants and environment are present and interact.

A similar result can be found by analyzing the level of physical activity and exercise chosen by the adolescents, as depicted in Figure 3. Again we see that the interaction between FTO and exercise is present, but it is not as pronounced as the interaction

[41] Reliable data on food intake are limited to the clinical visits that started at those ages.

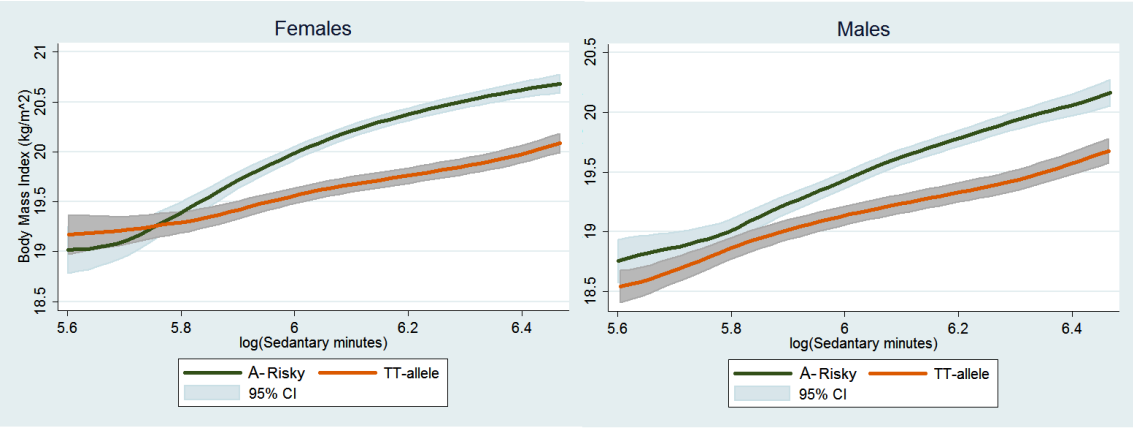
Figure 2: Gene-Food Interaction: the effect of caloric intake on BMI



Nonparametric local-mean smoothing using Epanechnikov kernel and Silverman's Rule-of-Thumb bandwidth. Combining information from successive clinical visits, age 10 to 14; excluding outliers in the top and bottom 5% of the distributions of BMI and log(food intake).

between FTO and food intake. More time spent in a sedentary lifestyle leads to an increase in BMI, but significantly more so for the children who happen to carry the risky A-allele. This is consistent with the predictions of the model: FTO induces a higher cost of dieting; this will induce a higher level of food intake; such intake will have an impact on fat-mass especially for those sedentary children who do not burn the excessive energy. Indeed the difference in BMI between the two genotypes cannot be detected at low levels of sedentary activity (a high investment in exercise).

Figure 3: Gene-Exercise Interaction: the effect of physical activity on BMI



Nonparametric local-mean smoothing using Epanechnikov kernel and Silverman's Rule-of-Thumb bandwidth. Combining information from successive clinical visits, age 10 to 14; excluding outliers in the top and bottom 5% of the distributions of BMI and log(sedentary minutes).

The robustness of these results is confirmed by the findings of other studies reporting evidence of the interaction between FTO and food intake, and FTO and exercise.[42] Also twin-studies support the finding that physical activity can attenuate the genetic determinants of obesity. Silventoinen et al. (2009) find that heritability of body-fat is higher among twins who are inactive, suggesting that "physical activity is able to modify the action of the genes responsible for predisposition to obesity." In other words, the evidence of this interaction is not limited to FTO but can extend to other parts of the genome.[43]

3.2.2 A Linear Production Function of Health

I consider a log-linear production function: $B_{t+1} = A(B_t, X)g^{\beta_g}\left[F_t^{\alpha_f(g)} \cdot E_t^{\alpha_e(g)}\right]$.[44] Following Ehrlich and Chuma (1990) and Galama et al. (2012a), I allow for decreasing returns to scale, so that the α_k do not need to sum to 1. The genetic-dependent parameter β_g captures differences in the constant productivity, equivalent to a Hicks-neutral technical change, and represents a vertical shift in figures (2) and (3). The parameters α_d and α_e capture changes in the marginal productivity of investments, equivalent to a non-neutral technical change, and represents shifts in the slopes, as suggested by the evidence of the previous section (3.2). The coefficients of these interactions between genetic variants g and investments (F_t, E_t) are used to test the existence of a productivity effect.

The term $A(X)$ controls for common variables that can influence the child's BMI, such as the past stock BMI B_{t-1}, and mother's health and socio-economic characteristics. The following log-linear equation is estimated:

$$log(B_{i,t}) = \mu + \beta_g g + \alpha_e log(E_{i,t}) + \alpha_f log(F_{i,t}) +$$

[42] Kilpeläinen et al. (2011) perform a meta-analysis of various studies in both adults and children; they find evidence of significant gene-lifestyle interaction for adults, but less prominent results in children and adolescents.

[43] See Qi and Cho (2008) and Qi et al. (2012) for some detailed examples.

[44] A similar specification of BMI as a function of caloric intake, exercise, and other observable characteristics is estimated by Rashad (2006), however she does not observe or model genetic differences. This functional form is commonly used in the literature, but does impose strong separability. Preliminary results estimating a CES function support this assumption, since I cannot reject that estimated elasticity of substitution is different from 1.

$$+ \alpha_{g\times e} log(E_{i,t}) \cdot g + \alpha_{g\times f} log(F_{i,t}) \cdot g + (1-\delta) log(B_{i,t-1})$$
$$+ \gamma_b log(B_i^{mom}) + h(X_{i,t}) + \kappa_t + \varepsilon_{i,t} \quad (2)$$

$B_{i,t}$ is the BMI of child i at time t; g is a dummy indicating the genotype of the child; $F_{i,t}$ are kilocalories consumed, as a measure of food consumption and investment in diet; $E_{i,t}$ are sedentary minutes, as a measure of investment in exercise; the coefficients $\alpha_{g\times f}$ and $\alpha_{g\times e}$ test for presence of a GxE, an investment specific productivity effect; β_g captures differences in the overall productivity due to the genetic variant; B_i^{mom} represents the mother's Body-Mass-Index before pregnancy; $h(X_{i,t})$ is a flexible function of control variables introduced to proxy for family and individual specific characteristics that might influence obesity and investment[45]; κ_t captures age effects. As a measure of genetic endowment g, I construct a dummy for whether the child is carrying at least one minor A-allele in the rs9939609 FTO gene variant. As a robustness check, I then consider a genetic-predisposition-score calculated as the number of obesity-related alleles of 24 different genes. The score is constructed following Speliotes et al. (2010)[46] and Vimaleswaran and Loos (2010). The obesity-genes were selected from the Genome-Wide Association Studies of Vimaleswaran and Loos (2010); Speliotes et al. (2010); Sandholt et al. (2012).[47] Mendel's law of independent assortment states that different genetic variants are uncorrelated, and indeed I find that all of these genetic loci have a very small correlation.[48] Consequently the genetic score displays a bell-shape similar to a normal distribution, as shown in figure 4. The underlying assumption in using this score is that each allele has the same marginal effect. In order to relax this assumption,

[45] I control for a polynomial of child's age at the clinic visit; the child's birth weight, as a proxy of prenatal investment; mother age at conception; dummies for different levels of mother and father Socio-Economic-Status and education levels; a dummy for teen-pregnancy; child parity

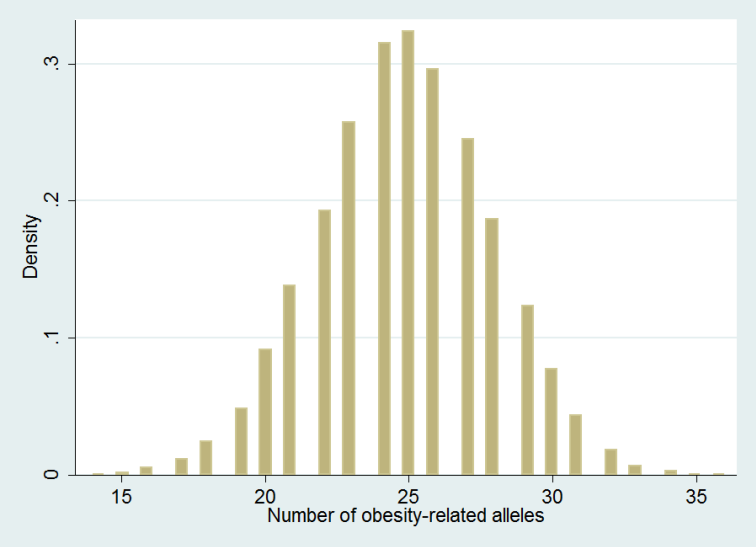
[46] They call it "genetic-susceptibility" score

[47] The genetic loci that I considered are: rs2229616 (variant of the MC4R gene), rs6548238 (TMEM18), rs9939609 (FTO), rs987237 (TFAP2B), rs7138803 (BCDIN3D), rs7647305 (ETV5), rs6265 (BDNF), rs10938397 (GNPDA2), rs1801282 (PPARG), rs7578597 (THADA), rs4402960 (IGF2BP2), rs12255372 (TCF7L2), rs1805081 (NPC1), rs10838738 (MTCH2), rs6235 (PCSK1), rs29941 (KCTD15), rs7498665 (SH2B1), rs10146997 (NRXN3), rs5015480 (HHEX), rs2605100 (LYPLAL1), rs1799884 (GCK), rs2815752 NEGR1), rs10508503 (PTER), rs780094 (GCKR). All of them are correlated to BMI in the ALSPAC sample, and have been validated in various studies as obesity-related genetic loci. For some there is evidence of potential environmental pathways through food intake (diet) or energy expenditure (exercise). See the discussion in section 2.

[48] All of the polychoric correlations are smaller than 0.05

and for comparability with the previous results, the genetic score g is dichotomized so that it is equal to one for the children who have more than the median number of obesity related alleles (number of 'fat-alleles' > 25).

Figure 4: Distribution of the Genetic-Predisposition-Score



Identification of the effect of the genetic variant rests on the assumption that genetic alleles are randomly inherited from the parents. This assumption is supported by Mendel's Law of Segregation, and is the foundation of the research on Mendelian Randomization initiated by Davey Smith (2003).

Table (3) reports the coefficients of equation (2) when using the risky A-allele of the FTO genetic variant as index of g.

As we can see from column (1) and (2), the genetic endowment of the child has a clear and strong effect on the obesity level, even after controlling for standard demographic characteristics as well as the mother BMI. The effect is similar to the ones found by related studies[49], and comparable to a 10% increase in the BMI of the mother. This is quite substantial, considering that obesity is a polygenic and complex disease and this is the effect of a single genetic variant. Controlling for the investment choices of the family in column (3), the coefficient β_g does not change. Once the interaction between the genetic variant and the two types of investment is introduced in column

[49]See Dina et al. (2007); Frayling et al. (2007); Timpson et al. (2008)

(4), the direct effect of FTO, β_g, still does not change and is equal to about a 1% change in BMI.[50] It is worth noticing the very high persistence of BMI: the past stock of BMI is a very important determinant of its current level of B_t, and its inclusion raises significantly the explanatory power. The most important contribution to BMI comes from food consumption and exercise, and their interaction with the genetic endowment. Most importantly, the interaction between the genetic variant and caloric intake is positive and statistically significant, as captured by the coefficient $\alpha_{g\times f}$. The individuals carrying the risky-generic variant convert calories into BMI at a rate $\approx 1/3$ higher. This coefficients captures a sizable shift in the slope and the productivity of investments in caloric intake.

Finally, robustness checks to the identification strategy are reported in columns (5) and (6), adding and removing relevant control variables in the spirit of Altonji et al. (2008).[51] According to Mendel's law of segregation, the offspring's genotype is random *conditional* on parental genotype. However only maternal but not paternal genotype is contained in the data. Flexibly controlling for maternal genotype does not change the coefficients of interest, as shown in column (5). Assuming that mother influence on the offspring behaviors are the most prominent and represent the greatest threat to identification, it can be envisioned that controlling for paternal genotype should not move the coefficients further. As an additional check to the robustness of the estimates, the estimated genetic effects are still very similar even when omitting the covariates X in column (6).

Tables (4-5) runs the model separately by gender, and performs various robustness checks. Gender differences appear in the relative importance of investments: girls' BMI is more influenced by food consumption, while for boys both food consumption and exercise play and important role. However the gender differences in coefficients are not statistically significant. So far I have considered BMI as a uniformly negative measure of health capital: the lower, the better. However this might not be an accurate characterization for those children with a very low level of body-mass. Less than 4%

[50] The genotype g is interacted with the de-meaned investment variables, $F_{i,t} - \bar{F}$ and $E_{i,t} - \bar{E}$, so that the coefficient β_g captures the effect at the average of the investment distribution.

[51] See also the discussion in Murphy and Topel (1990); Altonji et al. (2010); Oster (2014).

Table 3: Gene and Investment Interaction - FTO

		(1)	(2)	(3)	(4)	(5)	(6)
Risky FTO Gene	β_g	0.019	0.006	0.007	0.010	0.010	0.010
		[0.005]***	[0.002]***	[0.002]***	[0.002]***	[0.003]***	[0.003]***
log(Food Intake)	α_f			0.051	0.067	0.059	0.069
				[0.006]***	[0.009]***	[0.010]***	[0.009]***
G X Food Intake	$\alpha_{g\times f}$				0.025	0.027	0.026
					[0.011]**	[0.011]**	[0.011]**
log(Sedentary min.)	α_e			0.010	0.027	0.028	0.024
				[0.007]	[0.009]***	[0.011]***	[0.009]***
G X Sedentary min.	$\alpha_{g\times e}$				0.011	0.010	0.012
					[0.011]	[0.011]	[0.011]
$\log(BMI)_{t-1}$	$(1-\delta)$		0.969	0.939	0.939	0.947	0.967
			[0.007]***	[0.008]***	[0.008]***	[0.013]***	[0.008]***
$\log(BMI)_{mom}$	γ_b		0.090	0.090	0.090	0.097	
			[0.007]***	[0.007]***	[0.007]***	[0.012]***	
Covariates			X	X	X	X	
Mother Gene						X	
R^2		0.32%	78%	78%	78%	78%	78%
Observations		7052	7052	7052	7052	7052	7052

* significant at 10%; ** significant at 5%; *** significant at 1%. Standard error clustered at the individual level in brackets. Dependent variable: log BMI (kg/m^2); Risky FTO gene $g = 1$ if rs9939609 genetic variant contains one or more A-alleles; $g = 0$ otherwise; Covariates: gender; parity; age of child at clinic date; mom and dad education and SES; mother age at pregnancy; dummy for single mother; reliable dietary report; time dummy; month dummies; late respondent; birth weight. Mother gene: a dummy equal to one if the maternal rs9939609 genetic variant contains one or more A-alleles, interacted with log mother BMI, as well as the child's log caloric intake, log sedentary minutes and lagged log BMI. None of the interactions are statistically different from zero in column (5).

of the children are underweight in the sample considered, but the main results do not change when removing them from the estimating sample (see column 4). Therefore this does not seem to be a salient concern. Also changing the structure of the error term $\varepsilon_{i,t}$ does not change significantly the estimates; column 5 displays the estimates of a random effect model such that $\varepsilon_{i,t} = \mu_i + u_{i,t}$, where μ_i is a person-specific effect that is orthogonal to his genetic endowment and other observed characteristics. The results do change however when considering a fixed-effect model; this should not be surprising since genes are a "fixed" endowment of the individual: the relevant source of variation is not within-person and across time, but rather across children with different genotypes. Finally, the last three columns consider a different measurement of adiposity B_t as dependent variable. Column (7) estimates of the probability of being overweight[52]; column (8) estimates the effect of genetic variants and investment on changes in weight, controlling for height; and column (9) uses as dependent variable the age-and-sex standardized measurement of Body Mass Index (z-BMI score). All the estimated signs are in line with the main results, but for the absence of a significant interaction between diet and the FTO genotype in column (7), which limits the variation in the dependent variable to the probability of being overweight. This could be due to the fact that the FTO genetic variant has an influence on the whole distribution of BMI, not only the upper tail, as shown in figures (10-11) in the appendix.[53]

[52]Overweight is coded as 1 when BMI is higher than the sex-and-age adjusted 85% percentile of the 1990 British Growth Reference (see Cole et al. (1998)). About 23% of the sample is overweight.

[53]Further robustness checks are shown in the appendix. Tables (13) and (14) report the estimation of the linear health production separately by gender. Table (15) reports the estimates for the robustness checks, using the genetic score as a measure of g. Tables (16) and (17) report the estimation of the linear health production function using different measures of investments, both for food intake F_t and exercise E_e. The estimated productivity effects $\alpha_{g\times k}$ $k \in f, e$ do not change also when introducing an interaction between the genotype of the child and the lagged level of BMI $B_{i,t-1}$, or an interaction between g and the mother's BMI. Also introducing quadric terms for investments F_t and E_t does not influence the empirical conclusions. All results available upon request.

Table 4: Robustness Checks - FTO

		(1)	(2)	(3)	(4)	(5)	(6)
		Baseline	Males	Females	No Underweight	RE	FE
Risky FTO Gene	β_g	0.010	0.006	0.010	0.011	0.011	
		[0.002]***	[0.004]	[0.003]***	[0.003]***	[0.003]***	
log(Food Int.)	α_f	0.067	0.067	0.082	0.069	0.060	0.015
		[0.009]***	[0.013]***	[0.014]***	[0.009]***	[0.010]***	[0.012]
G X Food Int.	$\alpha_{g\times f}$	0.025	0.004	0.044	0.030	0.023	-0.004
		[0.011]**	[0.016]	[0.018]**	[0.011]***	[0.013]*	[0.015]
log(Sedentary m.)	α_e	0.027	0.042	0.007	0.028	0.040	0.026
		[0.009]***	[0.013]***	[0.013]	[0.009]***	[0.010]***	[0.012]**
G X Sedentary m.	$\alpha_{g\times e}$	0.012	0.026	-0.007	0.009	0.007	-0.016
		[0.011]	[0.016]*	[0.016]	[0.011]	[0.012]	[0.014]
B_{t-1}	$(1-\delta)$	0.939	0.947	0.928	0.911	0.815	-0.136
		[0.008]***	[0.012]***	[0.011]***	[0.008]***	[0.009]***	[0.017]***
Controls		X	X	X	X	X	X
R^2		78%	79%	79%	77%		64%
Observations		7,052	3,346	3,706	6,785	7,052	7,052

Column (1) reports the baseline estimates (same as table 3). Column (2) and (3) run the model separately for males and females. Column (4) runs the model dropping the children who are below the 5^{th} percentile of the z-BMI standard distribution for the UK (they represent 4% of the sample). Column (5) and (6) run the model using random effects and fixed effects, so that $\varepsilon_{i,t} = \mu_i + u_{i,t}$; all other columns report standard error clustered at the individual level. The dependent variable: log BMI (kg/m^2).

* significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors in brackets. Risky FTO gene $g = 1$ if rs9939609 genetic variant contains one or more A-alleles; $g = 0$ otherwise. Covariates: gender; parity; age of child at clinic date; mom and dad education and SES; mother age at pregnancy; dummy for single mother; reliable dietary report; time dummy; month dummies; late respondent; birth weight.

Table 5: Robustness Checks - FTO

		(1) Baseline	(2) Prob Overweight	(3) Weight	(4) zBMI	(5) Fat %
Risky FTO Gene	β_g	0.010	0.228	0.012	0.081	-0.011
		[0.002]***	[0.065]***	[0.003]***	[0.019]***	[0.019]
log(Food Int.)	α_f	0.067	0.500	0.072	0.490	0.036
		[0.009]***	[0.224]**	[0.011]***	[0.070]***	[0.078]
G X Food Int.	$\alpha_{g\times f}$	0.025	0.091	0.030	0.199	0.029
		[0.011]**	[0.274]	[0.013]**	[0.083]**	[0.093]
log(Sedentary m.)	α_e	0.027	0.554	0.031	0.189	0.141
		[0.009]***	[0.218]**	[0.011]***	[0.067]***	[0.068]**
G X Sedentary m.	$\alpha_{g\times e}$	0.012	0.082	0.009	0.076	0.021
		[0.011]	[0.252]	[0.013]	[0.080]	[0.081]
B_{t-1}	$(1-\delta)$	0.939	2.101	0.761	0.869	0.306
		[0.008]***	[0.052]***	[0.008]***	[0.008]***	[0.022]***
Controls		X	X	X	X	X
R^2		78%		88%	77%	55%
Observations		7,052	7,052	7,048	7,052	5,305

Column (1) reports the baseline estimates (same as table 3). Column (2) runs a probit model on the probability of being obese. Column (3) uses B_t =log(weight) as dependent variable, controlling for log(height). Column (4) uses z-BMI as dependent variable. Column (5) uses the estimated percentage of body fat as dependent variable. For all the other columns, the dependent variable: log BMI (kg/m^2).

* significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors in brackets. Risky FTO gene $g = 1$ if rs9939609 genetic variant contains one or more A-alleles; $g = 0$ otherwise. Covariates: gender; parity; age of child at clinic date; mom and dad education and SES; mother age at pregnancy; dummy for single mother; reliable dietary report; time dummy; month dummies; late respondent; birth weight.

Table (6) reports other estimates of the linear production function of health, this time using the genetic-obesity-score as index of g.

The main results carry through even when considering a polygenic approach. The genotype of the child is strongly related to his obesity, and both types of investments are important determinants of increases in BMI. There is evidence of interaction between diet and the genotype, since $\alpha_{g\times f}$ is positive and of similar magnitude; however there is no evidence of an interaction with investment in exercise, $\alpha_{g\times e} \approx 0$, even for males.[54] Finally, the magnitude of the estimated persistence of BMI, δ, is comparable to the estimates in table (3).

[54]For the robustness checks, see table (15) in the appendix.

Table 6: Gene and Investment Interaction - Genetic Score

		(1)	(2)	(3)	(4)	(5)
Risky Genetic	β_g	0.034	0.009	0.009	0.012	0.012
Score		[0.005]***	[0.002]***	[0.002]***	[0.002]***	[0.002]***
log(Food Intake)	α_f			0.051	0.065	0.066
				[0.006]***	[0.008]***	[0.008]***
G X Food Intake	$\alpha_{g\times f}$				0.025	0.026
					[0.011]**	[0.011]**
log(Sedentary min.)	α_e			0.010	0.019	0.014
				[0.007]	[0.008]**	[0.008]*
G X Sedentary min.	$\alpha_{g\times e}$				0.000	-0.003
					[0.011]	[0.011]
$\log(BMI)_{t-1}$	$(1-\delta)$		0.967	0.938	0.938	0.965
			[0.007]***	[0.008]***	[0.008]***	[0.008]***
$\log(BMI)_{mom}$	γ_b		0.089	0.090	0.090	
			[0.007]***	[0.007]***	[0.007]***	
Covariates			X	X	X	
R^2		1.05%	78%	78%	78%	78%
Observations		7052	7052	7052	7052	7052

* significant at 10%; ** significant at 5%; *** significant at 1%. Standard error clustered at the individual level in brackets. Dependent variable: log BMI (kg/m^2); Risky genetic score $g = 1$ if genetic score > 25; $g = 0$ otherwise; Covariates: gender; parity; age of child at clinic date; mom and dad education and SES; mother age at pregnancy; dummy for single mother; reliable dietary report; time dummy; month dummies; late respondent; birth weight.

3.2.3 Genetic Effects on the Investments

The theoretical analysis suggests that the genetic endowment of a child can have an effect not only on the production function of health $f(.;g)$, but also on the level of investments $I_t^* = (F_t^*, E_t^*)$. The optimal level of investment depends on the resources and the characteristics of the family, as well as from the genetic productivity effect $\partial f/\partial F_t$ and the genetic cost effect $\partial U/\partial F_t$. In order to test these predictions, I run a regression of investments $I_t = (F_t, E_t)$ on the genotype g of the child, lagged level of BMI $B_{i,t-1}$, and the usual control X.

$$\begin{aligned} \log(F_t) =& \mu_f + \beta_{g,f} g + \alpha_{B,f} log(B_{i,t-1}) + \beta_f X_{i,t} + \nu_{i,t}^f \\ \log(E_t) =& \mu_E + \beta_{g,E} g + \alpha_{B,E} log(B_{i,t-1}) + \beta_E X_{i,t} + \nu_{i,t}^E \end{aligned} \tag{3}$$

Table 7: Genetic Effect on Investments - FTO

	Caloric Consumption		**Sedentary Minutes**	
	Male	Female	Male	Female
	(1)	(2)	(3)	(4)
Risky FTO Gene	0.020	0.018	0.006	0.005
	[0.009]**	[0.008]**	[0.007]	[0.006]
$\log(BMI_{t-1})$	-0.103	-0.164	0.076	0.048
	[0.036]***	[0.032]***	[0.025]***	[0.021]**
Covariates	X	X	X	X
Observations	3,347	3,371	3,347	3,371

* significant at 10%; ** significant at 5%; *** significant at 1%. Standard error clustered at the individual level in brackets. Dependent variables: logarithm of daily kilocalories intake (columns (1) and (2)), and logarithm of daily sedentary minutes (columns (3) and (4)). Covariates: parity; age of child at clinic date; mom and dad education and SES; mother age at pregnancy; dummy for single mother; reliable dietary report; time dummy; month dummies; late respondent; birth weight.

The estimates of table (7) show a significant genetic difference in the level of food intake for both males and females. On the other side, FTO has no clear effect on the chosen minutes of sedentary activity. This is evidence in favor of the existence of a genetic cost effect for diet, but not for exercise. Since a high caloric intake is proportionally more risky for the children with the A-allele (the estimated coefficient $\hat{\alpha}_{g\times f} \geq 0$ in table (3)), we would expect them to eat less. Indeed, the first order

conditions show how F_t^* and $\alpha_d(g)$ are negatively related: a higher risk of obesity should induce a higher effort in dieting and lower caloric intake. On the other side, optimal caloric intake is directly related to preferences for food. Since we observe that children with the FTO A-allele tend to eat more, in spite of their increased risk, this suggests that $\frac{\partial U}{\partial F}\big|_{g=A} \geq \frac{\partial U}{\partial F}\big|_{g=T}$. In other words, there is evidence in favor of a genetic cost effect. This is consistent with the evidence from the molecular genetic literature, that shows how FTO influences food intake but not energy expenditure.

The results are somewhat different when considering the effect of the genetic score on investments. Table (8) shows evidence of a genetic cost effect for both diet and exercise, but only for females. Sedentary activity and caloric intake of boys do not seem to be affected by the composite measure of genetic predisposition to obesity. This can be due to the fact that the different genetic variants used to construct this genetic score do not work through the same biological pathways as FTO.

Table 8: Genetic Effect on Investments - Genetic Score

	Caloric Consumption		**Sedentary Minutes**	
	Male	Female	Male	Female
	(1)	(2)	(3)	(4)
Risky Genetic	0.011	0.014	0.001	0.022
Score	[0.009]	[0.008]*	[0.007]	[0.006]***
$\log(BMI_{t-1})$	-0.104	-0.166	0.076	0.041
	[0.036]***	[0.032]***	[0.025]***	[0.020]**
Covariates	X	X	X	X
Observations	3,347	3,371	3,347	3,371

* significant at 10%; ** significant at 5%; *** significant at 1%. Standard error clustered at the individual level in brackets. Dependent variables: logarithm of daily kilocalories intake (columns (1) and (2)), and logarithm of daily sedentary minutes (columns (3) and (4)). Covariates: parity; age of child at clinic date; mom and dad education and SES; mother age at pregnancy; dummy for single mother; reliable dietary report; time dummy; month dummies; late respondent; birth weight.

3.3 Replication using the FHS

It is imperative to replicate the analysis using a different dataset, in order to assure that the results presented so far are not an incidental fluke of the dataset used, but

that they uncover a strong biological connection between genetic endowment, healthy investment, and evolution of BMI.

For this purpose I use the offspring cohort of the Framingham Heart Study, a very different sample that follows adult individuals over various parts of the life-cycle, who have lived in the United States, and who were exposed to very different environments in terms of both accessibility to cheap-caloric food and chances to perform physical activity.

Table (9) mimics the results shown in table (3), estimating equation (2) with the FHS sample. Since the proxy of E_t used in this dataset is a self-reported measure of physical activity, we expect to find $\alpha_e \leq 0$. More physical activity should reduce fat-mass. Following the results of Rosenquist et al. (2014) who find a stronger association between body-mass-index and the FTO-genetic variant for the sample of adults born after the 1940's, I also split the sample according to date of birth in order to capture the potential effect of being exposed for a longer time period to an obesogenic environment. The first two columns of table (9) show that the FTO-genetic variant is associated to higher body-mass-index also in this sample, especially so for those born after 1940. The last two columns show evidence of a similar genetic productivity effect. The coefficient of the interaction between the risky-FTO genetic variant and caloric intake is always positive, and stronger if we consider only those born after 1940.

The fact that we obtain a higher coefficient for the younger generation is consistent with our model of the interaction between genes and investment. As shown in figures (5-6), younger individuals have been exposed since birth to an 'obesogenic' environment, where both the price of food and the share of income devoted to food consumption was declining, effectively reducing the cost of caloric intake. Such decline in cost becomes particularly salient for the individuals carrying the risky-variant of the FTO genotype.

While the evidence on the genetic productivity effect is replicated using the FHS dataset, the results on the genetic cost effect do not carry through. As shown in table (10), when evaluating the genetic cost effect I do not find any significant relation between carries of the risky A-allele and either food consumption or level of physical activity.

Table 9: FHS: Gene and Investment Interaction - FTO

		(1)	(2) born after 1940	(3)	(4) born after 1940
Risky FTO variant	β_g	0.024***	0.043***	0.002	0.005**
		[0.007]	[0.010]	[0.001]	[0.002]
log(Energy Intake)	α_f			0.013***	0.022***
				[0.004]	[0.005]
G X Energy Intake	$\alpha_{g\times f}$			0.010**	0.016**
				[0.005]	[0.006]
log(Physical Activity)	α_e			-0.005**	-0.009***
				[0.002]	[0.003]
G X Physical Activity	$\alpha_{g\times e}$			0.003	0.001
				[0.003]	[0.004]
$\log(BMI)_{t-1}$	$(1-\delta)$			0.937***	0.927***
				[0.006]	[0.009]
Covariates				x	x
R^2		0.4%	1.2%	85.3%	84.7%
Observations		8,258	4,642	8,258	4,642

Dependent variable: log BMI (kg/m^2); Risky FTO gene $g = 1$ if rs9939609 genetic variant contains one or more A-Alleles; $g = 0$ otherwise; Covariates: gender; 3-degree polynomial in age; dummies education and income; dummies for marital status; reliable dietary report; time dummies; birth cohort dummies; 20 first principal components of genome. Sample: Framingham Heart Study, Offspring Cohort individuals aged less than 70. * significant at 10%; ** significant at 5%; *** significant at 1%. Standard error clustered at the individual level in brackets.

Figure 5: Relative Food prices

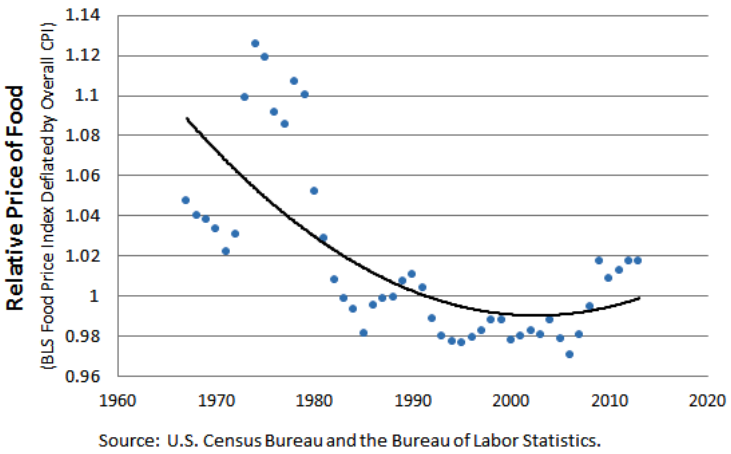
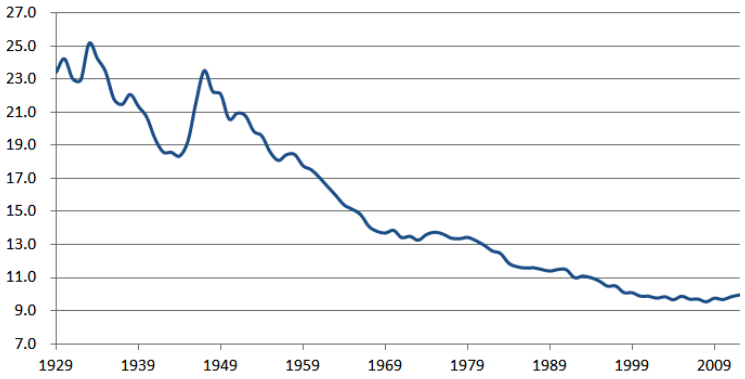


Figure 6: Food Expenditure as a Share of Disposable Income



Source: Economic Research Service, USDA; U.S. Census Bureau; and the Bureau of Labor Statistics.

Table 10: Cost Effect

	Caloric Consumption		**Physical Activity**	
	Whole sample	Born after 1940	Whole sample	Born after 1940
	(1)	(2)	(3)	(4)
Risky FTO	0.004	0.004	0.023	0.022
Gene	[0.011]	[0.016]	[0.013]	[0.019]
$\log(BMI_{t-1})$	0.000	0.035	-0.192***	-0.233***
	[0.033]	[0.046]	[0.035]	[0.046]
Covariates	X	X	X	X
Observations	8,258	4,642	8,258	4,642

Dependent variable: Dependent variables: log of daily kilocalories intake (columns (1) and (2)), and log of physical activity (columns (3) and (4)); Risky FTO gene $g = 1$ if rs9939609 genetic variant contains one or more A-Alleles; $g = 0$ otherwise; Covariates: $\log(BMI)_{t-1}$; log(physical activity) in columns (1) and (2), and log(kilocalories) in columns (3) and (4); gender; 3-degree polynomial in age; dummies education and income; dummies for marital status; reliable dietary report; time dummies; birth cohort dummies; 20 first principal components of genome. Sample: Framingham Heart Study, Offspring Cohort individuals aged less than 70. * significant at 10%; ** significant at 5%; *** significant at 1%. Standard error clustered at the individual level in brackets.

3.4 Calibration of the Dynamic Model

The analysis of the raw data gives suggestive evidence of gene-environment interaction, and the existence of both a genetic productivity effect and a genetic cost effect. I now turn to a parametric estimation of the dynamic model that allows me to take into consideration both types of investments at the same time, as well as controlling for some socio-economic characteristics that could influence the production function of health and BMI. Following the structure provided by the model, genes are not simply considered as an input into the production function, but rather they set the stage for the evolution of health and human capital.

In order to bring the model to the data, I estimate it recursively.[55]

Healthy Investment Production Function I assume a Cobb-Douglas for the health investment production function,

$$\begin{aligned} I(F_t, E_t; g) &= \log\left[\phi(g) F_t^{a(g)} E_t^{b(g)}\right] \\ &= \log\phi(g) + a(g)\log F_t + b(g)\log E_t \end{aligned}$$

so that the evolution of BMI over time follows the following law of motion:

$$\log B_{t+1} = \log\phi(g) + a(g)\log F_t + b(g)\log E_t + (1 - \delta_1 - t/T\delta_2)\log B_t + \varepsilon_t$$

The inverse of the production function with respect to the first input becomes:

$$\log F_t = \log\Im\left(B_{t+1} - (1-\delta_t)B_t - \varepsilon_t; E_t\right)$$

[55] Another option would be to estimate a version of the Euler Equation (11) as in Galama et al. (2012a) or Dustmann and Windmeijer (2000). With appropriate functional form assumption, the structural relation between health and investment become the following:

$$a_{1,t} I_t^{\frac{1}{a+b}-\alpha} + a_2 I_t^{\frac{1}{a+b}-\alpha}\frac{\Delta I_t}{I_t} = a_{3,t} B_t^{-1} + a_{4,t} B_t^{-(\gamma+1)} \tag{4}$$

where the time-varying coefficients depend non-linearly on the parameters of the model, wages, and prices.

$$= \frac{1}{a(g)} \left(\log B_{t+1} - (1 - \delta_t) \log B_t - \log \phi(g) - b(g) \log E_t - \log \varepsilon_t \right)$$

or in levels:

$$F_t = \Im \left(B_t, B_{t+1}, \varepsilon_t; E_t \right) = \left(\frac{B_{t+1}}{\phi(g)(1 - \delta_t) B_t \, E_t^{b(g)} \, \varepsilon_t} \right)^{1/a(g)}$$

Utility Function For simplicity, the utility function is also assumed to be a Cobb-Douglas:

$$u \left(B, F, \ell, c; g \right) = \zeta_B \log B + \zeta_F(g) \log F + \zeta_\ell \log \ell + \zeta_c \log c$$

where the preference for consumption is normalized to be $\zeta_c = 1 - \zeta_B - \zeta_F - \zeta_\ell$, and I allow the genetic endowment of the individual to change the preference parameter $\zeta_F(g)$, which governs the preferences regarding food consumption.

Bellman Equation Combining all the above functional forms, the Bellman equation becomes:

$$\begin{aligned} V_t(B_t, Y_t, \varepsilon_t; g) = \max_{E_t, B_{t+1}} \; & \zeta_B \log B_t + \frac{\zeta_F(g)}{a(g)} \left(\log B_{t+1} - (1 - \delta_t) \log B_t - \log \varepsilon_t \right) \\ & - \frac{b(g) \zeta_F(g)}{a(g)} \log E_t + \zeta_\ell \log \left(\Omega - E_t \right) + \zeta_c \log c \\ & + \beta E V_{t+1}(B_{t+1}, Y_{t+1}, \varepsilon_{t+1}; g) \end{aligned}$$

Optimal Time Allocation The optimal "static" allocation of time between leisure and exercise, as defined by equation (9), is the following:

$$\begin{aligned} \left[u_f \left(B, F^*, \Omega - E^*, c^*; g \right) - p_{F_t} u_c \left(B, F^*, \Omega - E^*, c^*; g \right) \right] \Im_E &= u_\ell \left(B, F^*, \Omega - E^* \right) \\ \zeta_F + \frac{a}{b} \frac{\zeta_\ell \, E_t^*}{\Omega - E_t^*} &= \frac{\zeta_c}{Y_t / \Im \left(B_t, B_{t+1}, \varepsilon_t; E_t \right) p_{F_t} - 1} \end{aligned}$$

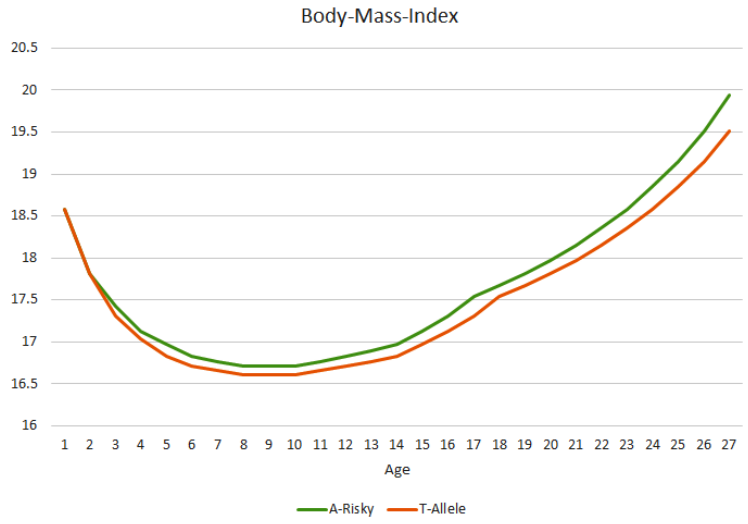
Euler Equation The first order condition of the value function maximization, as expressed in general terms in equation (11), take the following form:

$$[p_{F_t} u_c(B, F, \ell, c) - u_f(B, F, \ell, c)] \Im_{B'}(B, B', \varepsilon, E) =$$
$$\beta E [u_B(B', F', \ell') + u_F(B', F', \ell') \Im_{B'}(B', B'', \varepsilon', E')]$$
$$\left[p_{F_t} \frac{\zeta_c}{Y_t - p_{F_t} F_t} - \frac{\zeta_f}{F} \right] \frac{\Im}{a(g) B'} = \beta E \left[\frac{\zeta_B}{B'} - \frac{\zeta_f}{F'} \frac{(1-\delta')\Im'}{a(g) B'} \right]$$

where both consumption and exercise are expressed in terms the state variables, so that $F = \Im(B, B', \varepsilon, E) = \left[\frac{B' - (1-\delta)B - \varepsilon}{\phi E^{b(g)}} \right]^{1/a(g)}$ and $E = \hat{\xi}(B, B', \varepsilon)$ solving equation (9).

As a preliminary result, using the parameters estimated in the reduced form equations of the previous section I can calibrate the model to mimic the evolution of body mass index in the first years of the children's lives depicted in figure (1). As can be seen from the simulation in figure (7), although initially there is no genetic difference in the health capital of children, the differences in investment decisions driven by the genotype of the child start to cumulate over time, and therefore the simulated profiles of body mass index start diverging.

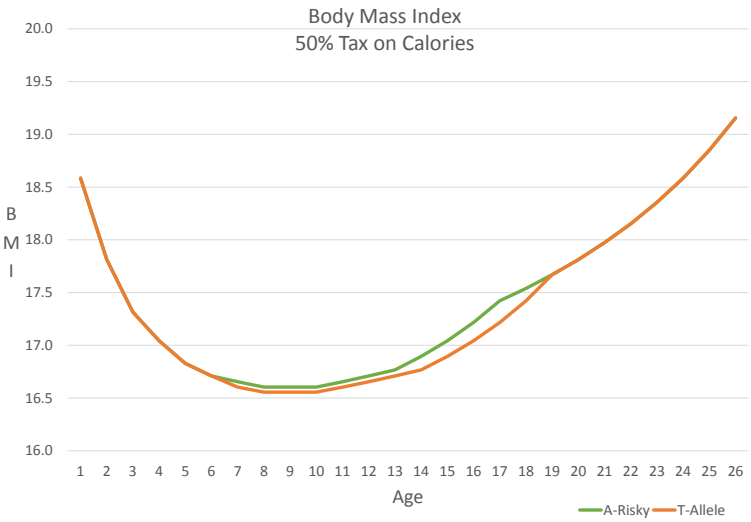
Figure 7: Simulated Evolution of Body Mass Index



Simulated evolution of BMI over time for children with a different genotype. Bellman Equation calibrated using the parameters estimated in the reduced form equations of sections (3.2.2-3.2.3).

Using this calibrated model, I can simulate the effect of various changes in the eco-

Figure 8: Policy A: Tax on Calories

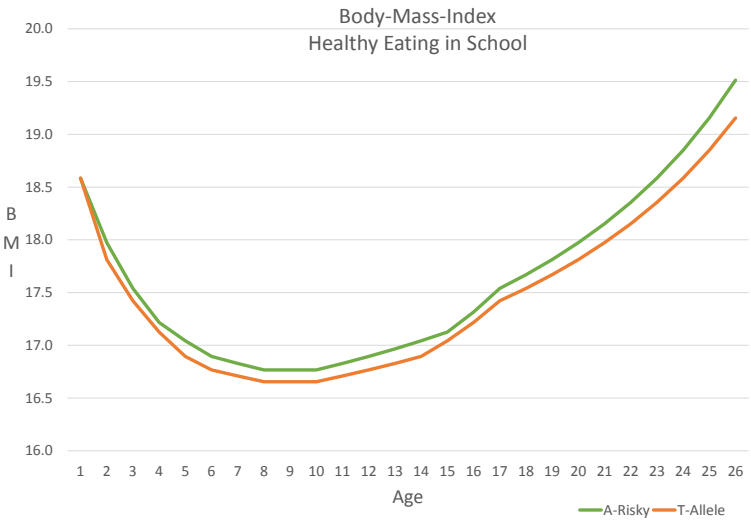


Simulated impact of a permanent 50% increase in price of food p_{F_t} on the life cycle evolution of BMI for children with a different genotype. Bellman Equation calibrated using the parameters estimated in the reduced form equations of sections (3.2.2-3.2.3).

nomic environment. In particular I consider two situations: one is an increase in the price of calories p_{F_t}, the other is a proportional decrease in the amount of caloric intake. The first counterfactual changes the cost of investments in diet, and investigates the differential reaction due to the genetic cost effect. The second counterfactual changes the chosen level of inputs in a proportional manner for both genotypes, therefore investigating the impact of the genetic productivity effect.

Figure (8) shows the effect on the evolution of BMI of a 50% increase in the price of calories. Such an increase could be due to a policy that directly taxes calories, or to a counterfactual economy with much higher food prices, similar to the situation faced by the early cohort of the Framingham Heart Study participants who were born before the 1940s. As expected, an increase in prices lowers caloric consumption and consequently BMI: the average reduction at the end of the period (beginning of adulthood) is about .6 BMI points (1.5kg, 3% reduction). Even more interestingly, such price changes induce different responses according to the genotype of the individual: carriers of the risky A-allele reduce caloric intake much more, so much that the overall genetic difference in BMI is negligible. This situation of non significant genetic differences in BMI replicates the findings from the older FHS cohort, born in a period with high relative food prices,

Figure 9: Policy B: Healthy Eating in School



Simulated impact of a permanent 25% reduction in optimal caloric intake F_t^* for the first 10 time periods $t = 0, ..., 10$ on the life cycle evolution of BMI for children with a different genotype. Bellman Equation calibrated using the parameters estimated in the reduced form equations of sections (3.2.2-3.2.3).

or the findings by Scannell Bryan et al. (2014), who look at genetic differences in BMI in rural Bangladesh: in both situations food was scarce and costly, and the genetic risk of overeating was not triggered because of economic constraints limiting the amount of food consumption. This simulation highlights the importance of taking into account the interaction between different genotypes and changes in the budget constraint.

The other counterfactual considered is a 25% reduction in food intake for the first 10 years of life. Such a policy can mimic the effect of an intervention that removes high-caloric foods from school cafeteria and replaces them with healthier options. As shown in figure (9), the effect of the policy on the evolution of BMI is more moderate, with an average reduction of .4 BMI points (.9 kg, 2% reduction). Contrary to before, this simulation forces a *proportional* decrease in caloric intake for both genotypes, and the genetic differences in BMI are still observable. However, such difference are reduced by 16%. This second policy shows that changes in BMI due to the genetic productivity effect are more moderate, but an equal reduction in inputs for both genotypes induces a more than proportional reduction in output. Furthermore, this counterfactual shows how changes limited to the first period of the life-cycle can have strong repercussion later in life: due to its strong persistence over time, the average BMI is lower than

before even after the policy is lifted and individuals can optimally choose their food consumption.

3.5 Decomposition of Total Genetic Effect

A simple decomposition can help understand the relative importance of the genetic cost effect and the genetic productivity effect. Consider the log of body mass index B as a linear function of different inputs $W = (F, E, X)$: $B = W'\beta$, where W are caloric intake F_t, exercise E_t, and other controls as in equation (2). Following Blinder (1973) and Oaxaca (1973), the average genetic difference in body mass, $\Delta B = E(B_A - B_T)$, can be decomposed in two components: the difference in parameters, and the difference in inputs:

$$E(B_A) - E(B_T) = \underbrace{[E(W_A) - E(W_T)]' \beta_A}_{\Delta\text{inputs}} + \underbrace{E(W_T)' (\beta_A - \beta_T)}_{\Delta\text{parameters}}$$

where $E(B_k)$ and $E(W_k)$ are the average values of BMI and the various inputs W for children with genotype k; β_k are the parameters of the production function of body mass estimated using only the children with genotype k.

In the sample the average difference in BMI attributable to the FTO genetic variant is $\Delta BMI = 0.37$. For the average child, who is 1.5m tall, this is roughly equivalent to an additional kilogram of weight. In terms of obesity rates, this genetic difference represents a 1.4-fold increase in the odds of being overweight. Albeit small, this difference is statistically and economically significant, considering the high cost of obesity calculated in Cawley (2010). Furthermore, studies in adult populations find even starker differences of about 3 additional kilograms or a 1.7-fold increase in the odds of obesity.[56] Using this decomposition, I find that differences in the parameters of the production function, $E(W_T)' (\beta_A - \beta_T)$, account for 35.4% of the average difference in fat mass (95% confidence interval (26%,39%)). In terms of the model, this means that the ge-

[56] Frayling et al. (2007). These small differences in outcomes are quite common in GWAS studies, especially when looking at a complex and polygenic trait such as fat mass. Furthermore consider that "FTO [...] remains the gene with the most robust association and greatest effect size." according to Yeo and O'Rahilly (2012).

netic productivity effect can explain less than half of the overall genetic difference in BMI. The more substantive part of this difference, 64.5% (with c.i. (47%,72%)), is attributable to changes in the optimal allocation of inputs: this is due to variation in the incentives that family face when investing in the human capital of the child.

In summary, I find that the FTO genetic variant tends to increase the average level of caloric intake of children, and at the same time it increases the risk of obesity for those who are abundant eaters. Coupling these two effects explains why children endowed with the risky A-allele of the FTO gene tend to be more obese.

4 Conclusion

Recent genetic research allows economists to observe and model some later factors which traditionally were treated as unobserved heterogeneity. The impact of innate ability on life-time wellbeing has long concerned both health and labor economists, since unobserved ability generates observed differences in outcomes. In order to solve this problem of endogenous differences, researchers often resort to econometric methods such as fixed-effects that control for unobserved heterogeneity among individuals and relegate these differences into latent variables that receive little direct attention.[57] However, some components of these latent traits can now be observed and modeled using recent developments in genetic research. Through an international scientific effort called the Human Genome Project, in 2003 researchers completed the full sequencing and mapping of all of the genes that constitute our DNA, opening the way for a better understanding of the biological functioning of the human DNA and the medical, social, and economic consequences of the differences embedded in our genetic code.[58]

I connect these recent findings in genetics with established models of human capital

[57] A common approach is to use twin-studies to control for genetic similarities among individuals. See the seminal work in economics of Taubman (1976), and the more recent work of Björklund and Jäntti (2012); Cesarini et al. (2009). Kohler et al. (2011) provide an excellent discussion on how to leverage twin-studies to model unobserved genetic endowments and causal pathways. Goldberger (1979); Molenaar (2010) provide critical analyses of these methods.

[58] See appendix A for a quick description of basic concepts in human genetics. See Lander (2011) for a discussion of the importance of the Human Genome Project and its consequences for our understanding of the biological functions encoded in the genome.

formation, and develop a model of investment in health and human capital that sheds light on the dynamic interaction between genetic types and investment choices. The genotype of an individual can shift preferences as well as human capital production possibility frontiers. Individual heterogeneity is modeled as differences in the parameters that govern the utility costs and the returns to different investments. In other words, I do not assume that DNA has a direct effect on an individual's outcomes, but rather I allow her genetic endowment to delineate the set of achievable combinations of choices (healthy behaviors) and outcomes (health) that she can attain. Therefore a genotype that has been related to a particular trait, such as obesity for instance,[59] confers a comparative advantage to attain a lean figure by changing the costs and the benefits of certain healthy behaviors. Though more costly, appropriate lifestyle choices over the years can compensate or accentuate genetic endowments.

Using this framework, I shed light on the mechanisms through which an individual's genotype shapes investments in her human capital, enriching our understanding of the causal mechanisms that connect genetic types to variations in outcomes.

Although the model can be generalized to different types of investments and different facets of human capital, in this paper I focus on obesity. A negative feature of health that increased alarmingly in the last decades, obesity is the second most important cause of premature death in the US. It has far-reaching economic and medical consequences, raging from lower work productivity to type 2 diabetes and coronary heart disease. An estimated $147 billion a year is spent in the US to cure obesity related illnesses.[60]

Obesity is well suited for this analysis because it has a strong genetic basis, yet it can be influenced by individual choices of diet and physical exercise. To measure endowed individual heterogeneity, I focus on a genetic variant of FTO, a gene associated to a 1.4-fold increase in the probability of being overweight for the 11 and 13-years-old adolescents of my sample, 60% of whom are carrier of the risky variant. Established findings in molecular genetics link FTO to hypothalamic regulation of appetite and food intake, suggesting that this genetic type increases the cost of following a strict

[59] See Sandholt et al. (2012); Speliotes et al. (2010) for a discussion of the genes that have been associated with obesity

[60] Finkelstein et al. (2009)

diet without altering the incentives to engage in physical activity.

This intuition is corroborated by the results estimated using two different datasets, the Avon Longitudinal Study of Parents And Children (ALSPAC), a very rich sample of British children born in 1991/1992, and the Offspring Cohort of the Framingham Heart Study (FHS), a sample of US adults followed since 1971. I find evidence that the FTO genotype changes both the production function of BMI and the level of healthy investment. Adolescents carrying the A-allele in the FTO genotype are at greater risk of obesity, even when consuming the same calories and performing the same physical activity as the adolescents with the other genotype. Their production function is shifted, with a rate of conversion of calories into BMI 1/3 steeper. Instead of counterbalancing this effect with a stricter diet, they tend to consume 2% more calories, suggesting a difference in their preferences for food intake. There is no evidence of a significant interaction between the FTO genotype and physical activity. These results are replicated when considering a genetic score for predisposition to obesity based on 26 different genetic variants, and the findings are robust to different specifications of the production function of BMI and the use of different measurements of obesity or investments. The qualitative results are also corroborated when looking at the FHS. Furthermore, I find that the estimated interaction between the FTO genotype and caloric intake is stronger for individuals born later, since they have been exposed since birth to an 'obesogenic' environment where calories are readily available and physical activity is costly and effortful.

I make use of these estimates in two ways, a decomposition exercise and the calibration of structural model. First, I decompose the difference in BMI across the FTO genotypes in two components: 35.4% of this differences is attributable to changes in productivity parameters of the BMI production function; the remaining 64.5% is attributable to changes in the investments. Secondly, calibrating a structural model, I evaluate the effect of two policies. A high calorie tax almost eliminates the difference in BMI across genetic groups. An early childhood policy target at reducing caloric intake in the first 10 years of life does not have the same equalizing effect, but still manages to lower BMI in the long run.

In future work, the model and the empirical analysis will be extended to consider the

effect of genotypes on habit formation, food addiction, or the process of learning one's type. Furthermore, a similar model can be used to evaluate the effect of genetic differences on other facets of human capital: both genes and environments have been proven to be powerful determinants of cognition, mental health, social skills, behavioral attitudes, many other facets of health, and even educational attainment.[61] Each outcome must be carefully associated with an appropriate choice of genotypes and investment.

On a more general level, recent results in the field of human genetics can be leveraged to improve our understanding of the inner mechanisms of human capital formation, and to shed light on the incentives that people face when making choices of investment in their skills. Integrating an economic framework of decision-making with the incremental findings of Genome-Wide-Association-Studies, we can estimate more reliably the causal effect of investment on human capital accumulation, we can understand how people differ in their response to incentives to invest in their skills, and finally we can properly tailor public interventions in health care, schooling, and education to the individual characteristics.

Acknowledgments

I am extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and Pietro Biroli will serve as guarantors for the contents of this paper

[61] See for example the results of Rietveld et al. (2013)

References

Adda, J. and Cooper, R. W. (2003). *Dynamic Economics: Quantitative Methods and Applications*. the MIT Press.

Altonji, J. G., Conley, T., Elder, T. E., and Taber, C. R. (2010). Methods for Using Selection on Observed Variables to Address Selection on Unobserved Variables. *Working Paper*, 1(September).

Altonji, J. G., Elder, T. E., and Taber, C. R. (2008). Using Selection on Observed Variables to Assess Bias from Unobservables when Evaluating Swan-Ganz Catheterization. *American Economic Review: Papers & Proceedings*, 98(2):345–350.

Beauchamp, J. P., Cesarini, D., van der Loos, M., Koellinger, P. D., Groenen, P. J. F., Fowler, J. H., Rosenquist, J. N., Thurik, A. R., and Christakis, N. A. (2011). Molecular Genetics and Economics. *Journal of Economic Perspectives*, 25(4):57–82.

Becker, G. S. (2007). *Economic Theory*. Transaction Publishers.

Beckers, S., Zegers, D., Van Gaal, L. F., and Van Hul, W. (2009). The role of the leptin-melanocortin signalling pathway in the control of food intake. *Critical reviews in eukaryotic gene expression*, 19(4):267–87.

Benjamin, D. J., Caplin, A., Cesarini, D., Thom, K., and Turley, P. (2014). Smoking, Genes, and Health. *Working Paper*, pages 1–27.

Benjamin, D. J., Chabris, C. F., Glaeser, E. L., Gudnason, V., Harris, T. B., Laibson, D. I., Launer, L. J., and Purcell, S. (2007). Genoeconomics. In Weinstein, M., Vaupel, J. W., and Wachter, K. W., editors, *Biosocial Surveys*, chapter 15, pages 304–335. The National Academies Press, Washington, D.C.

Björklund, A. and Jäntti, M. (2012). How important is family background for labor-economic outcomes? *Labour Economics*, 19(4):465–474.

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, 8(Autumn 4):436–455.

Bochukova, E. G., Huang, N., Keogh, J., Henning, E., Purmann, C., Blaszczyk, K., Saeed, S., Hamilton-Shield, J., Clayton-Smith, J., O'Rahilly, S., Hurles, M. E., and Farooqi, I. S. (2010). Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature*, 463(7281):666–70.

Boyd, A., Golding, J., Macleod, J., Lawlor, D. A., Fraser, A., Henderson, J., Molloy, L., Ness, A. R., Ring, S. M., and Davey Smith, G. (2013). Cohort Profile: The 'Children of the 90s'–the index offspring of the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology*, 42(1):111–127.

Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., Taylor, A., and Poulton, R. (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, 297(5582):851–4.

Cawley, J. (2010). The economics of childhood obesity. *Health affairs*, 29(3):364–71.

Cawley, J., Han, E., and Norton, E. C. (2011). The validity of genes related to neurotransmitters as instrumental variables. *Health Economics*, 20(May):884–888.

Cawley, J. and Meyerhoefer, C. (2012). The medical care costs of obesity: an instrumental variables approach. *Journal of Health Economics*, 31(1):219–30.

Cecil, J. E., Tavendale, R., Watt, P., Hetherington, M. M., and Palmer, C. N. A. (2008). An obesity-associated FTO gene variant and increased energy intake in children. *The New England Journal of Medicine*, 359:2558–2566.

Cesarini, D., Dawes, C. T., Johannesson, M., Lichtenstein, P., and Wallace, B. (2009). Genetic Variation in Preferences for Giving and Risk Taking. *Quarterly Journal of Economics*, 124(2):809–842.

Cole, T. J., Freeman, J. V., and Preece, M. A. (1998). British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Statistics in medicine*, 17(4):407–29.

Conley, D., Fletcher, J. M., and Dawes, C. T. (2014). The Emergence of SocioGenomics. *Contemporary Sociology: A Journal of Reviews*, 43(4):458–467.

Conti, G. and Heckman, J. J. (2010). Understanding the Early Origins of the Education-Health Gradient: A Framework That Can Also Be Applied to Analyze Gene-Environment Interactions. *Perspectives on Psychological Science*, 5(5):585–605.

Cunha, F. and Heckman, J. J. (2008). Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *The Journal of Human Resources*, 43:738–782.

Cutler, D. M., Glaeser, E. L., and Shapiro, J. M. (2003). Why Have Americans Become More Obese? *Journal of Economic Perspectives*, 17(3):93–118.

Daniels, S. R., Jacobson, M. S., McCrindle, B. W., Eckel, R. H., and Sanner, B. M. (2009). American Heart Association Childhood Obesity Research Summit: executive summary. *Circulation*, 119(15):2114–23.

Davey Smith, G. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*, 32(1):1–22.

Davey Smith, G. (2010). Mendelian Randomization for Strengthening Causal Inference in Observational Studies: Application to Gene x Environment Interactions. *Perspectives on Psychological Science*, 5(5):527–545.

Davey Smith, G. and Ebrahim, S. (2004). Mendelian randomization: prospects, potentials, and limitations. *International Journal of Epidemiology*, 33(1):30–42.

Dina, C., Meyre, D., Gallina, S., Durand, E., Körner, A., Jacobson, P., Carlsson, L. M. S., Kiess, W., Vatin, V., Lecoeur, C., Delplanque, J., Vaillant, E., Pattou, F., Ruiz, J., Weill, J., Levy-Marchal, C., Horber, F., Potoczna, N., Hercberg, S., Le Stunff, C., Bougnères, P., Kovacs, P., Marre, M., Balkau, B., Cauchi, S., Chèvre, J.-C., and Froguel, P. (2007). Variation in FTO contributes to childhood obesity and severe adult obesity. *Nature genetics*, 39(6):724–6.

Dustmann, C. and Windmeijer, F. (2000). Wages and the Demand for Health – A Life Cycle Analysis. *IZA Discussion Paper*, 171.

Ehrlich, I. and Chuma, H. (1990). A Model of the Demand for Longevity and the Value of Life Extension. *Journal of Political Economy*, 98(4):761–782.

Fang, M. Z. (2013). Violating the Monotonicity condition for instrumental variableDimorphic patterns of gene-behavior association. *Economics Letters*.

Fawcett, K. A. and Barroso, I. (2010). The genetics of obesity: FTO leads the way. *Trends in genetics*, 26(6):266–74.

Finkelstein, E. A., Graham, W. C. K., and Malhotra, R. (2014). Lifetime Direct Medical Costs of Childhood Obesity. *Pediatrics*, pages peds.2014–0063–.

Finkelstein, E. A., Trogdon, J. G., Cohen, J. W., and Dietz, W. (2009). Annual medical spending attributable to obesity: payer-and service-specific estimates. *Health affairs (Project Hope)*, 28(5):w822–31.

Fletcher, J. M. and Lehrer, S. F. (2011). Genetic lotteries within families. *Journal of Health Economics*, 30(4):647–59.

Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., Henderson, J., Macleod, J., Molloy, L., Ness, A. R., Ring, S. M., Nelson, S. M., and Lawlor, D. A. (2013). Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International Journal of Epidemiology*, 42(Feb):97–110.

Frayling, T. M., Timpson, N. J., Weedon, M. N., Zeggini, E., Freathy, R. M., Lindgren, C. M., Perry, J. R. B., Elliott, K. S., Lango, H., Rayner, N. W., Shields, B., Harries, L. W., Barrett, J. C., Ellard, S., Groves, C. J., Knight, B., Patch, A.-M., Ness, A. R., Ebrahim, S., Lawlor, D. A., Ring, S. M., Ben-Shlomo, Y., Jarvelin, M.-R., Sovio, U.,

Bennett, A. J., Melzer, D., Ferrucci, L., Loos, R. J. F., Barroso, I., Wareham, N. J., Karpe, F., Owen, K. R., Cardon, L. R., Walker, M., Hitman, G. A., Palmer, C. N. A., Doney, A. S. F., Morris, A. D., Smith, G. D., Hattersley, A. T., and McCarthy, M. I. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*, 316(5826):889–94.

Fredriksson, R., Hägglund, M., Olszewski, P. K., Stephansson, O., Jacobsson, J. A., Olszewska, A. M., Levine, A. S., Lindblom, J., and Schiöth, H. B. (2008). The obesity gene, FTO, is of ancient origin, up-regulated during food deprivation and expressed in neurons of feeding-related nuclei of the brain. *Endocrinology*, 149(5):2062–71.

Galama, T. (2015). A Contribution to Health Capital Theory. *CESR-Schaeffer Working Paper*, (004).

Galama, T., Hullegie, P., Meijer, E., and Outcault, S. (2012a). Is there Empirical evidence for decreasing returns to scale in a health capital model? *Health Economics*, 21:1080–1100.

Galama, T., Kapteyn, A., Fonseca, R., and Michaud, P.-C. (2012b). A Health Production Model with Endogenous Retirement. *Health Economics*, 902(August 2012):883–902.

Galama, T. and Van Kippersluis, H. (2010). A Theory of Socioeconomic Disparities in Health over the Life Cycle. *RAND Working Paper*, WR-773.

Galton, F. (1874). *English men of science: Their nature and nurture*. McMillan & Co, London.

Gobet, F. and Campitelli, G. (2007). The role of domain-specific practice, handedness, and starting age in chess. *Developmental psychology*, 43(1):159–72.

Goldberger, A. S. (1979). Heritability. *Economica*, 46(184):327–247.

Govaerts, C., Srinivasan, S., Shapiro, A., Zhang, S., Picard, F., Clement, K., Lubrano-Berthelier, C., and Vaisse, C. (2005). Obesity-associated mutations in the melanocortin 4 receptor provide novel insights into its function. *Peptides*, 26(10):1909–19.

Govindaraju, D. R., Cupples, L. A., Kannel, W. B., O'Donnell, C. J., Atwood, L. D., D'Agostino, R. B., Fox, C. S., Larson, M., Levy, D., Murabito, J. M., Vasan, R. S., Splansky, G. L., Wolf, P. A., and Benjamin, E. J. (2008). Genetics of the Framingham Heart Study population. *Advances in Genetics*, 62:33–65.

Gray, J., Yeo, G. S. H., Cox, J. J., Morton, J., Adlam, A.-L. R., Keogh, J. M., Yanovski, J. a., El Gharbawy, A., Han, J. C., Tung, Y.-C. L., Hodges, J. R., Raymond, F. L.,

O'rahilly, S., and Farooqi, I. S. (2006). Hyperphagia, severe obesity, impaired cognitive function, and hyperactivity associated with functional loss of one copy of the brain-derived neurotrophic factor (BDNF) gene. *Diabetes*, 55(12):3366–71.

Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy*, 80(2):223–255.

Grossman, M. (2000). The human capital model. *Handbook of Health Economics*, 1:347–408.

Heckman, J. J. (2007). The economics, technology, and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences*, 104(33):13250–5.

Hirsch, J. (2003). Obesity: Matter over Mind. *Cerebrum*.

Huszar, D., Lynch, C. A., Fairchild-Huntress, V., Dunmore, J. H., Fang, Q., Berkemeier, L. R., Gu, W., Kesterson, R. A., Boston, B. A., Cone, R. D., Smith, F. J., Campfield, A. L., Burn, P., and Lee, F. (1997). Targeted Disruption of the Melanocortin-4 Receptor Results in Obesity in Mice. *Cell*, 88(1):131–141.

Jou, C. (2014). The Biology and Genetics of Obesity – A Century of Inquiries. *New England Journal of Medicine*, 370(20):1874–1877.

Kilpeläinen, T. O., Qi, L., Brage, S., Sharp, S. J., Sonestedt, E., Demerath, E. W., Ahmad, T., Mora, S., Kaakinen, M., Sandholt, C. H., Holzapfel, C., Autenrieth, C. S., Hyppönen, E., Cauchi, S., He, M., Kutalik, Z., Kumari, M., Stančáková, A., Meidtner, K., Balkau, B., Tan, J. T., Mangino, M., Timpson, N. J., Song, Y., Zillikens, M. C., Jablonski, K. A., Garcia, M. E., Johansson, S., Bragg-Gresham, J. L., Wu, Y., van Vliet-Ostaptchouk, J. V., Onland-Moret, N. C., Zimmermann, E., Rivera, N. V., Tanaka, T., Stringham, H. M., Silbernagel, G., Kanoni, S., Feitosa, M. F., Snitker, S., Ruiz, J. R., Metter, J., Larrad, M. T. M., Atalay, M., Hakanen, M., Amin, N., Cavalcanti-Proença, C., Grøntved, A., Hallmans, G., Jansson, J.-O., Kuusisto, J., Kähönen, M., Lutsey, P. L., Nolan, J. J., Palla, L., Pedersen, O., Pérusse, L., Renström, F., Scott, R. A., Shungin, D., Sovio, U., Tammelin, T. H., Rönnemaa, T., Lakka, T. A., Uusitupa, M., Rios, M. S., Ferrucci, L., Bouchard, C., Meirhaeghe, A., Fu, M., Walker, M., Borecki, I. B., Dedoussis, G. V., Fritsche, A., Ohlsson, C., Boehnke, M., Bandinelli, S., van Duijn, C. M., Ebrahim, S., Lawlor, D. A., Gudnason, V., Harris, T. B., Sørensen, T. I. A., Mohlke, K. L., Hofman, A., Uitterlinden, A. G., Tuomilehto, J., Lehtimäki, T., Raitakari, O. T., Isomaa, B., Njølstad, P. R., Florez, J. C., Liu, S., Ness, A., Spector, T. D., Tai, E. S., Froguel, P., Boeing, H., Laakso, M., Marmot, M., Bergmann, S., Power, C., Khaw, K.-T., Chasman, D., Ridker, P., Hansen, T., Monda, K. L., Illig, T., Järvelin, M.-R., Wareham, N. J., Hu, F. B., Groop, L. C., Orho-Melander, M., Ekelund, U., Franks, P. W., and Loos, R. J. F. (2011). Physical activity attenuates the influence of FTO variants on obesity risk: a meta-analysis of 218,166 adults and 19,268 children. *PLoS medicine*, 8(11):e1001116.

Kohler, H.-P., Behrman, J. R., and Schnittker, J. (2011). Social Science Methods for Twins Data: Integrating Causality, Endowments, and Heritability. *Biodemography and Social Biology*, 57(1):88–141.

Lakdawalla, D. and Philipson, T. J. (2009). The growth of obesity and technological change. *Economics and Human Biology*, 7(3):283–93.

Lakdawalla, D., Philipson, T. J., and Bhattacharya, J. (2005). Welfare-Enhancing Technological Change and the Growth of Obesity. *American Economic Review*, 95(2):253–257.

Lander, E. S. (2011). Initial impact of the sequencing of the human genome. *Nature*, 470(7333):187–97.

LeBlanc, E. S., O'Connor, E., Whitlock, E., Patnode, C., and Kapka, T. (2011). Effectiveness of primary carerelevant treatments for obesity in adults: a systematic evidence review for the US Preventive Services Task Force. *Annals of internal Medicine*, 155:434–447.

Li, Z., Zhou, Y., Carter-Su, C., Myers, M. G., and Rui, L. (2007). SH2B1 enhances leptin signaling by both Janus kinase 2 Tyr813 phosphorylation-dependent and -independent mechanisms. *Molecular endocrinology (Baltimore, Md.)*, 21(9):2270–81.

Lundborg, P. and Stenberg, A. (2010). Nature, nurture and socioeconomic policy-what can we learn from molecular genetics? *Economics and Human Biology*, 8(3):320–30.

Mattocks, C., Ness, A. R., Leary, S. D., Tilling, K., Blair, S. N., Shield, J., Deere, K., Saunders, J., Kirkby, J., Davey Smith, G., Wells, J. C., Wareham, N. J., Reilly, J. J., and Riddoch, C. J. (2008). Use of accelerometers in a large field-based study of children: protocols, design issues, and effects on precision. *Journal of physical activity & health*, 5 Suppl 1:S98–111.

Moffitt, T. E., Caspi, A., and Rutter, M. (2005). Strategy for investigating interactions between measured genes and measured environments. *Archives of general psychiatry*, 62(5):473–81.

Molenaar, P. C. M. (2010). On the Limits of standard quantitative genetic modeling of inter-individual variation: extensions, ergodic conditions and a new genetic factor model of intra-individual variation. In Hood, K. E., Halpern, C. T., Greenberg, G., and Lerner, R. M., editors, *Handbook of Developmental Science, Behavior, and Genetics*, chapter 19. Wiley-Blackwell, Oxford, UK.

Mulcaster, R. (1582). *Mulcaster's Elementaire*. Clarendon Press, London.

Murphy, K. M. and Topel, R. H. (1990). Efficiency Wages Reconsidered: Theory and Evidence. In Weiss, Y. and Fishelson, S., editors, *Advances in the Theory and Measurement of Unemployment*, pages 204–249. St. Martin's Press, New York, NY.

Norton, E. C. and Han, E. (2008). Genetic information, obesity, and labor market outcomes. *Health economics*, 17(9):1089–104.

Oaxaca, R. L. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(October 3):693–709.

Ogden, C. L., Carroll, M. D., Kit, B. K., and Flegal, K. M. (2012). Prevalence of obesity and trends in body mass index among US children and adolescents, 1999-2010. *JAMA*, 307(5):483–90.

Ogden, C. L., Carroll, M. D., Kit, B. K., and Flegal, K. M. (2014). Prevalence of Childhood and Adult Obesity in the United States, 2011-2012. *JAMA*, 311(8):806.

Ogden, C. L., Flegal, K. M., Carroll, M. D., and Johnson, C. L. (2002). Prevalence and Trends in Overweight Among US Children and Adolescents, 1999-2000. *JAMA: The Journal of the American Medical Association*, 288(14):1728–1732.

Olszewski, P. K., Fredriksson, R., Olszewska, A. M., Stephansson, O., Alsiö, J., Radomska, K. J., Levine, A. S., and Schiöth, H. B. (2009). Hypothalamic FTO is associated with the regulation of energy intake not feeding reward. *BMC neuroscience*, 10:129.

Oster, E. (2014). Unobservable Selection and Coefficient Stability: Theory and Evidence. *Mimeo, Yale University*.

Plomin, R., Defries, J. C., Craig, I. W., and Mcguffin, P. (2008). *Behavioral Genetics*. Worth Publishers.

Plomin, R., DeFries, J. C., and Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological bulletin*, 84(2):309–22.

Price, G. M., Paul, A. A., Key, F. B., Harter, A., Cole, T. J., Day, K. C., and Wadsworth, M. E. J. (1995). Measurement of diet in a large national survey: comparison of computerized and manual coding of records in household measures. *Journal of Human Nutrition and Dietetics*, 8(6):417–428.

Purcell, S. (2002). Variance Components Models for GeneEnvironment Interaction in Twin Analysis. *Twin Research and Human Genetics*, 5(06):554–571.

Qi, L. and Cho, Y. A. (2008). Gene-environment interaction and obesity. *Nutrition reviews*, 66(12):684–94.

Qi, L., Kraft, P., Hunter, D. J., and Hu, F. B. (2008). The common obesity variant near MC4R gene is associated with higher intakes of total energy and dietary fat, weight change and diabetes risk in women. *Human molecular genetics*, 17(22):3502–8.

Qi, Q., Li, Y., Chomistek, A. K., Kang, J. H., Curhan, G. C., Pasquale, L. R., Willett, W. C., Rimm, E. B., Hu, F. B., and Qi, L. (2012). Television watching, leisure time physical activity, and the genetic predisposition in relation to body mass index in women and men. *Circulation*, 126(15):1821–7.

Rashad, I. (2006). Structural estimation of caloric intake, exercise, smoking, and obesity. *The Quarterly Review of Economics and Finance*, 46(2):268–283.

Ren, D., Zhou, Y., Morris, D., Li, M., Li, Z., and Rui, L. (2007). Neuronal SH2B1 is essential for controlling energy and glucose homeostasis. *The Journal of Clinical Investigation*, 117(2):397–406.

Rietveld, C. A., Conley, D., Eriksson, N., Esko, T., Medland, S. E., Vinkhuyzen, A. A. E., Yang, J., Boardman, J. D., Chabris, C. F., Dawes, C. T., Domingue, B. W., Hinds, D. A., Johannesson, M., Kiefer, A. K., Laibson, D. I., Magnusson, P. K. E., Mountain, J. L., Oskarsson, S., Rostapshova, O., Teumer, A., Tung, J., Visscher, P. M., Benjamin, D. J., Cesarini, D., and Koellinger, P. D. (2014a). Replicability and Robustness of Genome-Wide-Association Studies for Behavioral Traits. *Psychological Science*, 25(11):1975–1986.

Rietveld, C. A., Esko, T., Davies, G., Pers, T. H., Turley, P., Benyamin, B., Chabris, C. F., Emilsson, V., Johnson, A. D., Lee, J. J., de Leeuw, C., Marioni, R. E., Medland, S. E., Miller, M. B., Rostapshova, O., van der Lee, S. J., Vinkhuyzen, A. A. E., Amin, N., Conley, D., Derringer, J., van Duijn, C. M., Fehrmann, R., Franke, L., Glaeser, E. L., Hansell, N. K., Hayward, C., Iacono, W. G., Ibrahim-Verbaas, C., Jaddoe, V. W. V., Karjalainen, J., Laibson, D. I., Lichtenstein, P., Liewald, D. C., Magnusson, P. K. E., Martin, N. G., McGue, M., McMahon, G., Pedersen, N. L., Pinker, S., Porteous, D. J., Posthuma, D., Rivadeneira, F., Smith, B. H., Starr, J. M., Tiemeier, H., Timpson, N. J., Trzaskowski, M., Uitterlinden, A. G., Verhulst, F. C., Ward, M. E., Wright, M. J., Davey Smith, G., Deary, I. J., Johannesson, M., Plomin, R., Visscher, P. M., Benjamin, D. J., Cesarini, D., and Koellinger, P. D. (2014b). Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences*, 111(38):13790–13794.

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., Westra, H.-J., Shakhbazov, K., Abdellaoui, A., Agrawal, A., Albrecht, E., Alizadeh, B. Z., Amin, N., Barnard, J., Baumeister, S. E., Benke, K. S., Bielak, L. F., Boatman, J. A., Boyle, P. A., Davies, G., de Leeuw, C., Eklund, N., Evans, D. S., Ferhmann, R., Fischer, K., Gieger, C., Gjessing, H. K., Hägg, S., Harris, J. R., Hayward, C., Holzapfel, C., Verbaas, C. A. I., Ingelsson, E., Jacobsson, B., Joshi, P. K., Jugessur, A., Kaakinen, M., Kanoni, S., Karjalainen, J., Kolcic, I., Kristiansson, K., Kutalik, Z., Lahti, J., Lee, S. H., Lin, P., Lind, P. A., Liu, Y., Lohman, K., Loitfelder, M., McMahon, G., Vidal, P. M., Meirelles, O., Milani, L., Myhre, R., Nuotio, M.-L., Oldmeadow, C. J., Petrovic, K. E., Peyrot, W. J., Polašek, O., Quaye, L., Reinmaa,

E., Rice, J. P., Rizzi, T. S., Schmidt, H., Schmidt, R., Smith, A. V., Smith, J. A., Tanaka, T., Terracciano, A., van der Loos, M. J., Vitart, V., Völzke, H., Wellmann, J., Yu, L., Zhao, W., Allik, J., Attia, J. R., Bandinelli, S., Bastardot, F., Beauchamp, J. P., Bennett, D. A., Berger, K., Bierut, L. J., Boomsma, D. I., Bültmann, U., Campbell, H., Chabris, C. F., Cherkas, L., Chung, M. K., Cucca, F., de Andrade, M., Jager, P. L. D., Neve, J.-E. D., Deary, I. J., Dedoussis, G. V., Deloukas, P., Dimitriou, M., Eiriksdottir, G., Elderson, M. F., Eriksson, J. G., Evans, D. M., Faul, J. D., Ferrucci, L., Garcia, M. E., Grönberg, H., Gudnason, V., Hall, P., Harris, J. M., Harris, T. B., Hastie, N. D., Heath, A. C., Hernandez, D. G., Hoffmann, W., Hofman, A., Holle, R., Holliday, E. G., Hottenga, J.-J., Iacono, W. G., Illig, T., Järvelin, M.-R., Kähönen, M., Kaprio, J., Kirkpatrick, R. M., Kowgier, M., Latvala, A., Launer, L. J., Lawlor, D. A., Lehtimäki, T., Li, J., Lichtenstein, P., Lichtner, P., Liewald, D. C., Madden, P. A., Magnusson, P. K., Mäkinen, T. E., Masala, M., McGue, M., Metspalu, A., Mielck, A., Miller, M. B., Montgomery, G. W., Mukherjee, S., Nyholt, D. R., Oostra, B. A., Palmer, L. J., Palotie, A., Penninx, B., Perola, M., Peyser, P. A., Preisig, M., Räikkönen, K., Raitakari, O. T., Realo, A., Ring, S. M., Ripatti, S., Rivadeneira, F., Rudan, I., Rustichini, A., Salomaa, V., Sarin, A.-P., Schlessinger, D., Scott, R. J., Snieder, H., Pourcain, B. S., Starr, J. M., Sul, J. H., Surakka, I., Svento, R., Teumer, A., Study, T. L. C., Tiemeier, H., van Rooij, F. J., Wagoner, D. R. V., Vartiainen, E., Viikari, J., Vollenweider, P., Vonk, J. M., Waeber, G., Weir, D. R., Wichmann, H.-E., Widen, E., Willemsen, G., Wilson, J. F., Wright, A. F., Conley, D., Davey Smith, G., Franke, L., Groenen, P. J. F., Hofman, A., Johannesson, M., Kardia, S. L., Krueger, R. F., Laibson, D. I., Martin, N. G., Meyer, M. N., Posthuma, D., Thurik, A. R., Timpson, N. J., Uitterlinden, A. G., van Duijn, C. M., Visscher, P. M., Benjamin, D. J., Cesarini, D., and Koellinger, P. D. (2013). GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment. *Science*, 25(4):57–82.

Rosenquist, J. N., Lehrer, S. F., Malley, A. J. O., Zaslavsky, A. M., Smoller, J. W., and Christakis, N. A. (2014). Cohort of birth modifies the association between FTO genotype and BMI. *Proceedings of the National Academy of Sciences*.

Rutter, M. (2006). *Genes and Behavior: Nature-Nurture Interplay Explained*. Blackwell Publishers, Oxford, UK.

Rzehak, P., Scherag, A., Grallert, H., Sausenthaler, S., Koletzko, S., Bauer, C. P., Schaaf, B., von Berg, A., Berdel, D., Borte, M., Herbarth, O., Krämer, U., Illig, T., Wichmann, H.-E., Hebebrand, J., and Heinrich, J. (2010). Associations between BMI and the FTO gene are age dependent: results from the GINI and LISA birth cohort studies up to age 6 years. *Obesity facts*, 3(3):173–80.

Sandholt, C. H., Hansen, T., and Pedersen, O. (2012). Beyond the fourth wave of genome-wide obesity association studies. *Nutrition and Diabetes*, 2(7):e37.

Scannell Bryan, M., Argos, M., Pierce, B., Tong, L., Rakibuz-Zaman, M., Ahmed, A., Rahman, M., Islam, T., Yunus, M., Parvez, F., Roy, S., Jasmine, F., Baron, J. A., Kibriya, M. G., and Ahsan, H. (2014). Genome-Wide Association Studies and Heritability Estimates of Body Mass Index Related Phenotypes in Bangladeshi Adults. *PLoS ONE*, 9(8):e105062.

Scholz, J. K. and Seshadri, A. (2013). Health and Wealth In a Lifecycle Model. *Working Paper*, (July).

Silventoinen, K., Hasselbalch, A. L., Lallukka, T., Bogl, L., Pietila, K. H., Heitmann, B. L., Schousboe, K., Rissanen, A., Kyvik, K. O., Sørensen, T. I. A., and Kaprio, J. (2009). Modification effects of physical activity and protein intake on heritability of body size and composition. *American Journal of Clinical Nutrition*, 90:1096–1103.

Singh, A. S., Mulder, C., Twisk, J. W. R., van Mechelen, W., and Chinapaw, M. J. M. (2008). Tracking of childhood overweight into adulthood: a systematic review of the literature. *Obesity Reviews*, 9(5):474–88.

Speakman, J. R., Rance, K. A., and Johnstone, A. M. (2008). Polymorphisms of the FTO gene are associated with variation in energy intake, but not energy expenditure. *Obesity*, 16(8):1961–5.

Speliotes, E. K., Willer, C. J., Berndt, S. I., Monda, K. L., Thorleifsson, G., Jackson, A. U., Lango Allen, H., Lindgren, C. M., Luan, J., Mägi, R., Randall, J. C., Vedantam, S., Winkler, T. W., Qi, L., Workalemahu, T., Heid, I. M., Steinthorsdottir, V., Stringham, H. M., Weedon, M. N., Wheeler, E., Wood, A. R., Stefansson, K., North, K. E., McCarthy, M. I., Hirschhorn, J. N., Ingelsson, E., and Loos, R. J. F. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature genetics*, 42(11):937–48.

Stamatakis, E., Zaninotto, P., Falaschetti, E., Mindell, J., and Head, J. (2010). Time trends in childhood and adolescent obesity in England from 1995 to 2007 and projections of prevalence to 2015. *Journal of epidemiology and community health*, 64(2):167–74.

Taubman, P. (1976). Earnings, education, genetics, and environment. *Journal of Human Resources*, 11(4):447–461.

Taylor, A. E., Ebrahim, S., Ben-Shlomo, Y., Martin, R. M., Whincup, P. H., Yarnell, J. W., Wannamethee, S. G., and Lawlor, D. A. (2010). Comparison of the associations of body mass index and measures of central adiposity and fat mass with coronary heart disease, diabetes, and all-cause mortality: a study using data from 4 UK cohorts. *The American journal of clinical nutrition*, 91(3):547–56.

Timpson, N. J., Emmett, P. M., Frayling, T. M., Rogers, I. S., Hattersley, A. T., McCarthy, M. I., and Davey Smith, G. (2008). The fat mass- and obesity-associated

locus and dietary intake in children. *The American Journal of Clinical Nutrition*, 88(4):971–8.

Tung, Y.-C. L., Ayuso, E., Shan, X., Bosch, F., O'Rahilly, S., Coll, A. P., and Yeo, G. S. H. (2010). Hypothalamic-specific manipulation of Fto, the ortholog of the human obesity gene FTO, affects food intake in rats. *PloS one*, 5(1):e8771.

Tung, Y.-C. L. and Yeo, G. S. H. (2011). From GWAS to biology: lessons from FTO. *Annals of the New York Academy of Sciences*, 1220:162–71.

Unger, T. J., Calderon, G. a., Bradley, L. C., Sena-Esteves, M., and Rios, M. (2007). Selective deletion of Bdnf in the ventromedial and dorsomedial hypothalamus of adult mice results in hyperphagic behavior and obesity. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(52):14265–74.

Valette, M., Bellisle, F., Carette, C., Poitou, C., Dubern, B., Paradis, G., Hercberg, S., Muzard, L., Clément, K., and Czernichow, S. (2012). Eating behaviour in obese patients with melanocortin-4 receptor mutations: a literature review. *International Journal of Obesity*, 13(September):1–9.

Vimaleswaran, K. S. and Loos, R. J. F. (2010). Progress in the genetics of common obesity and type 2 diabetes. *Expert reviews in molecular medicine*, 12(February):e7.

von Hinke Kessler Scholder, S., Davey Smith, G., Lawlor, D. A., Propper, C., and Windmeijer, F. (2011). Mendelian randomization: the use of genes in instrumental variable analyses. *Health Economics*, 20(May):893–896.

von Hinke Kessler Scholder, S., Davey Smith, G., Lawlor, D. A., Propper, C., and Windmeijer, F. (2013). Child height, health and human capital: Evidence using genetic markers. *European Economic Review*, 57:1–22.

Wabitsch, M., Moss, A., and Kromeyer-Hauschild, K. (2014). Unexpected plateauing of childhood obesity rates in developed countries. *BMC medicine*, 12:17.

Ward, M. E., McMahon, G., St Pourcain, B., Evans, D. M., Rietveld, C. A., Benjamin, D. J., Koellinger, P. D., Cesarini, D., Davey Smith, G., and Timpson, N. J. (2014). Genetic variation associated with differential educational attainment in adults has anticipated associations with school performance in children. *PloS one*, 9(7):e100248.

Wardle, J., Carnell, S., Haworth, C. M. A., Farooqi, I. S., O'Rahilly, S., and Plomin, R. (2008). Obesity associated genetic variation in FTO is associated with diminished satiety. *The Journal of clinical endocrinology and metabolism*, 93(9):3640–3.

Whitlock, G., Lewington, S., Sherliker, P., Clarke, R., Emberson, J., Halsey, J., Qizilbash, N., Collins, R., and Peto, R. (2009). Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. *Lancet*, 373(9669):1083–96.

Yang, W., Kelly, T., and He, J. (2007). Genetic epidemiology of obesity. *Epidemiologic reviews*, 29(1):49–61.

Yeo, G. S. H. and O'Rahilly, S. (2012). Uncovering the biology of FTO. *Molecular Metabolism*, pages 32–36.

Appendix to "Genetic and Economic Interaction in Health Formation: The Case of Obesity."

September 10, 2015

Latest version here

A Basic concepts of human genetics

Our DNA contains between 20,000 and 25,000 genes which form the blueprint for the development and the functioning of the human body. Genes are the coding regions of DNA that are scattered over the human genome. The nucleus of every human cell contains the same DNA code, which was inherited at conception from the parents. Each parent contributes to one copy of the 23 pairs of chromosomes that make up the total of human DNA, so that each genetic locus contains two alleles, one coming from the father and the other from the mother. Since each sperm or egg cell contains only one of these alleles, a child has a 50% chance of receiving a particular allele from a particular parent. For example, suppose that at a given locus there are two possible alleles, A and b. If both parents have one A and one b allele at that locus (genotype 'Ab' - "heterozygous"), then every child has a 25% chance of being "homozygous" for A ('AA'), a 50% chance of being heterozygous ('Ab'), and a 25% chance of being homozygous for b ('bb').

While most of the DNA sequence is the same from one person to the next, so that (almost) all humans are born with 5 fingers and two eyes, there are some parts of the DNA that differ across people. The most common form of genetic variation are SNPs, Single Neuclotide Polymorphisms, and they represent a difference in a single DNA building block (nucleotide). There are roughly 10 million SNPs in the human genome. A genotype is the genetic type of a particular person at a precise genetic locus. A phenotype is the observable trait or outcome of interest which can be influenced by a particular genotype.

Genes transcribe proteins, which in turn regulate bodily structure and function through a cascade of biological processes. The observable outcomes of interest, the phenotypes, are far downstream this chain of causation that originates from the genotype. While in some cases one single genetic variant can directly lead to a disease (such as Sickle Cell Anemia or Huntington disease), most of the phenotypes are polygenic in nature, meaning that they are influenced by a multiple genes (Plomin et al. (2008)).

B The Bellman Equation

Using a recursive formulation, the model can be written using the following Bellman equation:

$$
\begin{aligned}
V_t(B_t, Y_t, \varepsilon_t; g) &= \max_{E_t, NF_t} U\left(B_t, F_t, \ell_t, c_t; g\right) + \beta E V_{t+1}(B_{t+1}, Y_{t+1}, \varepsilon_{t+1}; g) \\
s.t & \\
\Omega(B_t) &= \ell_t + E_t \qquad (5) \\
Y_t &= p_{F_t} F(UF_t, NF_t) + c_t \qquad (6) \\
F_t &= \kappa UF_t + NF_t \qquad (7) \\
B_{t+1} &= I(F_t, E_t; g) + (1 - \delta_t) B_t + \varepsilon_t \qquad (8)
\end{aligned}
$$

where total caloric consumption F_t, the time spent exercising E_t, and non-food consumption c_t are the control variables. Variability in the model is capture by ε_t, representing a shock in the production function of body-mass-index.

The total amount of time available depends on the health of the individual $\Omega(B_t) = \ell_t + E_t$, so that leisure becomes $\ell_t = \Omega(B_t) - E_t$. For simplicity I assume that BMI has no significant effect on time available, so that $\Omega(B_t) = \Omega, \forall B$.[62] Finally, Y_t is an exogenous inflow of income each period.

Using the budget set (6), we can express non-food consumption c_t as a function of food intake:

$$c_t = Y_t - p_{F_t} F(UF_t, NF_t)$$

Finally, using the law of motion of health (8), we can express calories F_t as a function of exercise E_t and current and future health (B_t, B_{t+1}):

$$
\begin{aligned}
B_{t+1} &= I(F_t, E_t) + (1 - \delta_t) B_t + \varepsilon_t \\
I(F_t, E_t) &= B_{t+1} - (1 - \delta_t) B_t - \varepsilon_t \\
\Rightarrow F_t &= \Im\left(B_t, B_{t+1}, \varepsilon_t, E_t\right)
\end{aligned}
$$

where $\Im$ is the inverse of the health investment function with respect to its first argument, and its existence is ensured by the assumption of monotonicity of the health investment function $I(.)$.

Therefore the bellman equation becomes:

$$V_t(B_t, Y_t, \varepsilon_t; g) = \max_{E_t, B_{t+1}} u\left(B_t, \Im\left(B_t, B_{t+1}, \varepsilon_t, E_t\right), \Omega - E_t, Y - p_f \Im\left(B_t, B_{t+1}, \varepsilon_t, E_t\right); g\right)$$

[62] The $\Omega(B_t)$ representation, that follows Grossman (1972), captures the idea that as weight increases it becomes harder to exercise. A similar effect can be found by allowing an interaction in the utility function between exercise and health. Note that if leisure is not valued, then all time available will be used to invest in health.

$$+ \beta EV_{t+1}(B_{t+1}, Y_{t+1}, \varepsilon_{t+1}; g)$$

The solution procedure will follow in two steps: first, given a set of state variables today and tomorrow and the realized shock (B, B', ε), the optimal choice of exercise E will be determined as a "static" choice; secondly, the we will solve the dynamic choice regarding the optimal stock of health B', taking as given the optimal allocation of time chosen previously.[63]

1. For a given (B, B', ε), define the optimal allocation of time $E^* = \hat{\xi}(B, B', \varepsilon)$ as the solution to the following static problem:

$$\varphi(B, B', \varepsilon;) = \max_E u(B, \Im(B, B', \varepsilon, E), \Omega - E, Y - p_f \Im(B, B', \varepsilon, E); g) \quad (9)$$

   The first order condition require that the optimal time allocation satisfies the following:

$$E^* : \quad [u_f(B, F^*, \Omega - E^*, c^*; g) - p_f u_c(B, F^*, \Omega - E^*, c^*; g)] \Im_E = u_\ell(B, F^*, \Omega - E^*)$$

   where $F^* = \Im(B, B', \varepsilon, E^*)$, $c^* = Y - p_f F^*$ and $\Im_E = \partial \Im(B, B', \varepsilon, E) / \partial E$

2. Once this optimal allocation is found, we can substitute this into the bellman equation and express everything in terms of the current state variables (B, ε) and the choice of the future health stock B':

$$V(B, Y, \varepsilon; g) = \max_{B'} \varphi(B, B', \varepsilon; g) + \beta EV'(B', Y', \varepsilon'; g) \quad (10)$$

   The first order condition to this problem is:

$$\varphi_{B'}(B, B', \varepsilon; g) + \beta EV'_{B'}(B', Y', \varepsilon'; g) = 0$$
$$\varphi_{B'}(B, B', \varepsilon; g) + \beta E\varphi_{B'}(B', B'', \varepsilon'; g) = 0$$

   where the second line is obtained by taking the derivative of (10) with respect to its first argument B and considering the equality one period ahead, since it must hold for every period. Expressing these derivatives in using (9), we have that the first order conditions are:

$$[p_f u_c(B, F, \ell, c) - u_f(B, F, \ell, c)] \Im_{B'}(B, B', \varepsilon, E) = \quad (11)$$
$$\beta E [u_{B'}(B', F', \ell', c') + u_{F'}(B', F', \ell', c') \Im_{B'}(B', B'', \varepsilon', E')]$$

[63] See (Adda and Cooper, 2003, ch 5.4)and the references therein. For notational simplicity I omit income (Y, Y_{t+1}) and genes g since they are not changing over time. For simplicity of notation, we use X to indicate current variables, and X' to indicate future variables.

C Data and Measurements Used

C.1 Avon Longitudinal Study of Parents and Children (ALSPAC)

An extremely rich dataset collected by epidemiologic researchers from the University of Bristol, the ALSPAC follows prospectively a cohort of children born from mothers living in a health district in the former County of Avon, in the South West of England, with an expected delivery date between April 1991 and December 1992.[64] The children from 14,541 pregnancies were initially recruited.

Detailed information has been collected since pregnancy using self-administered questionnaires, data extraction from research clinics and medical notes, linkage to routine information systems. This dataset contains state-of-the-art measures of cognition, behavior, genetics and health of the child as well as indicators of personality, physical and mental wellbeing of both the mother and the father (when present). The existence of a full set of biomarkers and genetic data constitutes an unprecedented chance to rigorously estimate the health of the child and the existence of gene-environment interaction into a comprehensive framework of skill formation within the family.

Please note that the study website contains details of all the data that is available through a fully searchable data dictionary, http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/

Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees

The following measures have been used in the analysis:

Anthropometric measures

Height was measured by using a Harpenden stadiometer (Holtain Ltd, Crymych, United Kingdom), and weight was assessed by using a weighing scale (Tanita TBF 305; Tanita UK Ltd, Yewsley, United Kingdom). A Lunar Prodigy DXA scanner (GE Medical Systems Lunar, Madison, WI) provided measures of body composition, including fat, lean body mass, and bone mass. Body mass index (BMI = weight (kg)/height squared (m^2)), and BMI normal z-scores were calculated from the 1990 British Growth Reference.[65] Although multiple measure of obesity are provided, BMI is used throughout the paper because it was measured most frequently, it is easily comparable to many other studies, and it provides an easy yet reliable measure of obesity risk.[66]

Food Consumption

Three-day dietary records including 2 weekday and 1 weekend day were obtained from adolescents a few days before the clinic visit; parents provided assistance as needed. Participants were instructed to record all foods and beverages consumed by using standard household measures. Records were reviewed during clinic visits to improve com-

[64] See Fraser et al. (2013); Boyd et al. (2013)

[65] See Cole et al. (1998)

[66] See Taylor et al. (2010) for a discussion of the reliability of BMI in predicting coronary heart disease, diabetes, and all-cause mortality, as compared to other measures of adiposity.

pleteness. Questionnaires queried for information on vitamin supplements, type of milk or fat spreads consumed, and details of other foods commonly eaten. Diet records were coded and analyzed by using the Diet In Data Out software (MRC Human Nutrition Research, Elsie Widdowson Laboratory, Cambridge, United Kingdom), which generates food codes and weights of each item recorded (Price et al. (1995)). Average daily nutrient intakes were calculated by using BRIGADE (University of Bristol, Bristol, United Kingdom) - a nutrient analysis program based on a nutrient databank that included the fifth edition of McCance and Widdowson's food tables and supplements. Nutrients for foods not in the databank were obtained from the National Diet and Nutrition Survey nutrient databases or calculated from the manufacturer's label.

Physical activity

The Actigraph uni-axial accelerometer (Actigraph, Fort Walton Beach, FL) was used to measure physical activity and has been validated for use in children and adolescents (Mattocks et al. (2008)). The accelerometer, which is worn around the waist, captures the frequency and intensity of movement in the vertical plane. Adolescents were asked to wear the accelerometer for 7 days during waking hours and to remove the instrument only during showering, bathing, and swimming. Physical activity measured directly from accelerometers (not including time spent swimming or cycling) was used. The accelerometers used in this study measured 1-min epochs. Adolescents with more than 3 days of accelerometer data were included in the analyses.

C.2 Framingham Heart Study (FHS)

The Framingham Heart Study (FHS) was initiated in 1948, when approximately 5,200 men and women 30-60 years of age residing in Framingham, Massachusetts were enrolled in the original cohort. The FHS was initiated to prospectively investigate risk factors for cardiovascular disease. Since then it has come to be composed of four separate but related cohort populations. Nearly six decades later, the FHS is the longest running, multigenerational longitudinal study in medical history.[67]

Framingham Offspring Study The Framingham Offspring Study began in 1971, consisting of 5,124 individuals who represented the children of the original cohort population and their spouses. The FHS study is remarkable in that there has been almost no loss to follow-up in the study. Participants in the offspring cohort were given physical examinations and detailed questionnaires at regular intervals starting in 1972, with a total of 8 waves completed through 2008. BMI was calculated from recorded height and weight. Notably, the offspring cohort was born over a 40-year period, with participants ranging in age from their teens to their late 50's at the time of study onset in 1971. IRB approval has been consistently maintained over the course of the study, and all subjects were consented by the FHS. In addition to providing survey and examination data, a large proportion of participants (73.0%, 3,742 individuals) had their DNA genotyped

[67] Govindaraju et al. (2008)

using Affymetrix SNP chips. Genotypes at the rs9939609 allele were extracted using the Plink software from data contained in the Framingham SHARe database.

D Summary Statistics and Distributions

D.1 Summary Statistics

The tables below report the summary statistics of different variables used in the model.

Table (11) reports the average measure of anthropometrics, food intake, and physical activity. It can be seen that the main results of mean differences across FTO-genotype are similar regardless of what is the particular measurement used. Height was measured by using a Harpenden stadiometer (Holtain Ltd, Crymych, United Kingdom), and weight was assessed by using a weighing scale (Tanita TBF 305; Tanita UK Ltd, Yewsley, United Kingdom). A Lunar Prodigy DXA scanner (GE Medical Systems Lunar, Madison, WI) provided measures of body composition, including fat, lean body mass, and bone mass. Body mass index (BMI = weight (kg)/height squared (m^2)), and BMI normal z-scores were calculated from the 1990 British Growth Reference.[68]. Three-day dietary records including 2 weekday and 1 weekend day were obtained from adolescents a few days before the clinic visit; parents provided assistance as needed. Participants were instructed to record all foods and beverages consumed by using standard household measures. Records were reviewed during clinic visits to improve completeness. Questionnaires queried for information on vitamin supplements, type of milk or fat spreads consumed, and details of other foods commonly eaten. Diet records were coded and analyzed by using the Diet In Data Out software (MRC Human Nutrition Research, Elsie Widdowson Laboratory, Cambridge, United Kingdom), which generates food codes and weights of each item recorded (Price et al. (1995)). Average daily nutrient intakes were calculated by using BRIGADE (University of Bristol, Bristol, United Kingdom) - a nutrient analysis program based on a nutrient databank that included the fifth edition of McCance and Widdowson's food tables and supplements. Nutrients for foods not in the databank were obtained from the National Diet and Nutrition Survey nutrient databases or calculated from the manufacturer's label. Food groups were formed on the basis of nutrient composition and culinary use of foods consumed. Dairy and milk groups were categorized into full-fat, low-fat, and nonfat on the basis of fat content. Total milk intake included full-fat, low-fat and nonfat plain and flavored milk. Total dairy intake included milk, cheese, cream, and yogurt; butter was not included. The Actigraph uni-axial accelerometer (Actigraph, Fort Walton Beach, FL) was used to measure physical activity and has been validated for use in children and adolescents (Mattocks et al. (2008)). Variables derived from the Actigraph were counts per minute as an estimate of total activity, minutes of sedentary activity, and minutes of moderate-to-vigorous activity (MVPA). On the basis of the results from a calibration study (Mattocks et al. (2008)), daily minutes of MVPA were defined by

[68] See Cole et al. (1998)

using cutoffs developed for moderate activity (accelerometer output between 3600 and 6200 counts/min) and vigorous activity (more than 6200 counts/min); time spent performing MVPA were summed to quantify minutes of MVPA. Self-reported physical activity was the answer to the question "In the past month, what was the average number of times that you participated in vigorous physical activity (such as running, dance, gymnastics,netball, swimming, or aerobics)?", with the answers being 1= none, 2 = less than once a week, 3 = 1-3 times a week, 4 = 4-6 times a week, 5 = daily.

Table(12) reports the average value of the covariates used in the regression tables, split by genotype of the child. It can be seen that only mother's BMI changes with FTO (the mother's and the child's genotype are correlated, and FTO is related to adiposity in both generations). The group mean of all the other variables are not statistically different from each other when the sample is split according to the FTO-gene. Mother and father education are reported on a scale from 1 to 5, from lowest to highest, where 1 is Certificate of Secondary Education (CSE) or less; 2 is a Vocational school; 3 is Ordinary-level of high school; 4 is Advanced-level of high school; 5 is a post-secondary Degree. Mother and Father Socio-Economic-Status (SES) are reported on a scale from 1 to 6, from highest to lowest; they are derived from self-reported occupation using the OPCS job codes, so that 1 is a professional worker, while 6 is an unskilled worker.

Table 11: Summary Statistics

	FTO gene type		
	T-allele	A-Risky	Total
Height	154.51	154.81	154.7
(cm)	[0.21]	[0.16]	[0.13]
Weight	46.22***	47.26***	46.88
(kg)	[0.24]	[0.18]	[0.14]
BMI	19.10***	19.47***	19.33
kg/cm^2	[0.07]	[0.05]	[0.04]
BMI z-score	0.20***	0.35***	0.3
	[0.02]	[0.02]	[0.01]
Fat Percentage	24.31***	25.42***	25.02
	[0.19]	[0.15]	[0.12]
Overweight (%)	22.17***	28.49***	26.19
	[0.82]	[0.67]	[0.52]
Underweight (%)	4.18	3.56	3.79
	[0.40]	[0.28]	[0.23]
Arm Circ.	23.90***	24.34***	24.18
(cm)	[0.07]	[0.05]	[0.04]
Waist Circ.	68.45***	69.39***	69.05
(cm)	[0.19]	[0.14]	[0.11]
Waist/Hip ratio	0.82	0.82	0.82
	[0.00]	[0.00]	[0.00]
Kilocalories	1.89**	1.92**	1.91
(x1000)	[0.01]	[0.01]	[0.01]
Fat Intake	75.82**	77.10**	76.64
(grams/day)	[0.45]	[0.33]	[0.27]
Dietary Cholesterol Intake	188.66**	193.39**	191.67
(grams/day)	[1.88]	[1.44]	[1.15]
Carbohydrate Intake	252.83*	255.58*	254.58
(grams/day)	[1.30]	[0.94]	[0.76]
Total Sugar Intake	114.74	115.87	115.46
(grams/day)	[0.91]	[0.64]	[0.53]
Physical Activity	7.51	7.55	7.54
(Sedentary Hours)	[0.02]	[0.02]	[0.01]
Physical Activity	23.92	23.68	23.77
(Moderate To Vigorous)	[0.32]	[0.25]	[0.20]
Physical Activity	581.96	576.78	578.66
(counts per minute)	[3.73]	[2.84]	[2.26]
Very Active	3.69	3.71	3.7
(self-report)	[0.02]	[0.01]	[0.01]

Average measures of adiposity, investment in diet, and investment in exercise. Pooled across gender and ages, separated by FTO-genotype. Standard errors of means in brackets. Mean difference * significant at 10%; ** significant at 5%; *** significant at 1%. Body mass index normal z-scores calculated using 1990 British Growth Reference. Fat percentage: ratio of fat mass to total mass. Overweight and Underweight calculated using the BMI z-scores with a cutoff of 5% and 85%. 3-day dietary records coded using the Diet In Data Out software. Actigraph data: counts per min., min. of sedentary activity, and moderate to vigorous activity. Self-reported activity ranged from 1 (never) to 5 (daily).

Table 12: Control Variables, by Child FTO genotype

	FTO gene type		
	T-allele	A-Risky	Total
Mother Edu	3.36	3.33	3.34
	[0.03]	[0.02]	[0.02]
Father Edu	3.32	3.34	3.33
	[0.04]	[0.03]	[0.02]
Mother SES	2.75	2.78	2.77
	[0.02]	[0.02]	[0.02]
Father SES	2.88	2.84	2.86
	[0.03]	[0.02]	[0.02]
Mother BMI	22.74**	23.00**	22.90
	[0.10]	[0.08]	[0.06]
Mother age	29.33	29.35	29.34
at birth	[0.12]	[0.09]	[0.07]
Teen mother (%)	1.51	2.10	1.88
	[0.33]	[0.29]	[0.22]
Single Mother (%)	15.85	15.28	15.49
	[0.98]	[0.73]	[0.58]
Parity	0.69	0.73	0.72
	[0.02]	[0.02]	[0.01]
Birth Weight (kg)	3.42	3.43	3.42
	[0.01]	[0.01]	[0.01]

Average value of the covariates for the sample used in the main analysis. Pooled across genders and separated by FTO-genotype. Standard errors of means in brackets. Mean difference * significant at 10%; ** significant at 5%; *** significant at 1%.

Education ranges from lowest (1 = CSE or less) to highest (5 = degree). Socio-Economic-Status ranges from from highest (1 = professional) to lowest (6 = unskilled). Teen mother is a dummy for mothers who were pregnant before age 19. Single mother is a dummy for a household without a male figure.

Figure 10: Distribution of BMI, Females

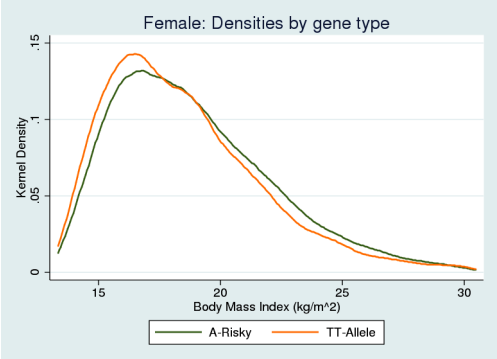
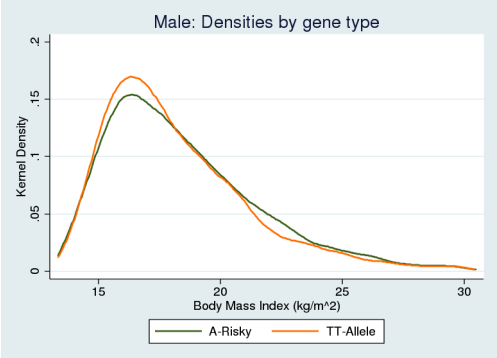


Figure 11: Distribution of BMI, Males



D.2 Distributions

Here below are the empirical distributions of the relevant variables, divided by gender and genetic endowment g. Figures (10) and (11) display the distribution of Body Mass Index, H; figures (12) and (13) display the distribution of the investment in diet, caloric intake C_t; figures (14) and (15) display the distribution of the investment in exercise, E_t;

E Robustness Checks

In this appendix I report some robustness checks of the main estimation results.

Tables (13) and (14) report the estimation of the linear health production separately by gender. Table (15) reports the estimates for the robustness checks, using the genetic score as a measure of g. Tables (16) and (17) report the estimation of the linear health production function using different measures of investments, both for diet C_t and exercise E_t.

Figure 12: Distribution of Diet, Females

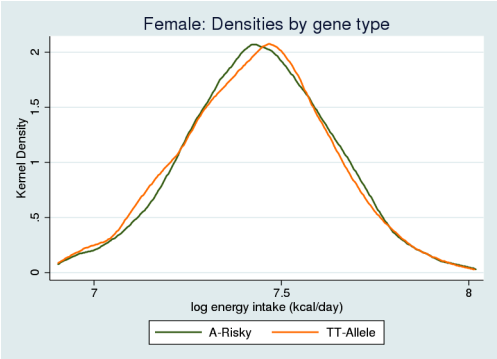


Figure 13: Distribution of Diet, Males

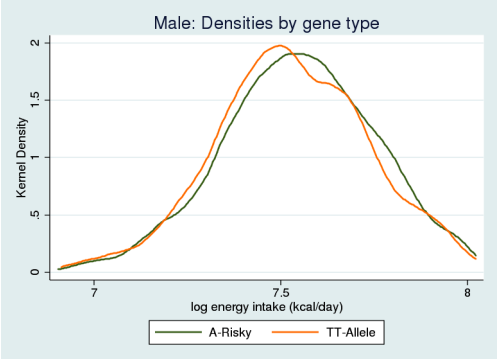


Figure 14: Distribution of Activity, Females

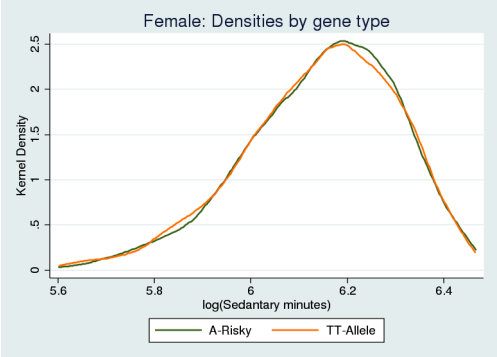


Figure 15: Distribution of Activity, Males

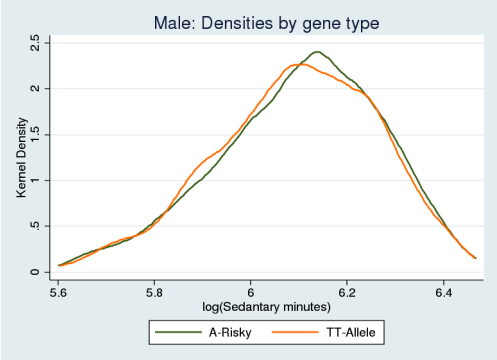


Table 13: Gene and Investment Interaction - FTO, Females

		(1)	(2)	(3)	(4)	(5)
Risky FTO Gene	β_g	0.022	0.005	0.008	0.010	0.012
		[0.007]***	[0.003]**	[0.003]**	[0.003]***	[0.003]***
log(Food Intake)	α_f			0.053	0.082	0.082
				[0.009]***	[0.014]***	[0.014]***
G X Food Intake	$\alpha_{g\times f}$				0.044	0.047
					[0.018]**	[0.018]***
log(Sedentary minutes)	α_e			0.010	0.007	0.005
				[0.007]	[0.013]	[0.013]
G X Sedentary min.	$\alpha_{g\times e}$				-0.007	-0.002
					[0.016]	[0.016]
log(BMI_{t-1})	$(1-\delta)$		0.948	0.928	0.928	0.961
			[0.010]***	[0.011]***	[0.011]***	[0.011]***
log(BMI)$_{mom}$	γ_b		0.105	0.104	0.104	
			[0.010]***	[0.010]***	[0.010]***	
Covariates			X	X	X	
R^2		0.41%	78%	78%	79%	77%
Observations		3706	3706	3706	3706	3706

* significant at 10%; ** significant at 5%; *** significant at 1%. Standard error clustered at the individual level in brackets. Dependent variable: log BMI (kg/m^2); Covariates: parity; age of child at clinic date; mom and dad education and SES; mother age at pregnancy; dummy for single mother; reliable dietary report; time dummy; month dummies; late respondent; birth weight.

Table 14: Gene and Investment Interaction - FTO, Males

		(1)	(2)	(3)	(4)	(5)
Risky FTO Gene	β_g	0.017	0.007	0.007	0.006	0.006
		[0.007]**	[0.003]***	[0.003]**	[0.004]	[0.005]
log(Food Intake)	α_f			0.064	0.067	0.069
				[0.008]***	[0.013]***	[0.013]***
G X Food Intake	$\alpha_{g\times f}$				0.004	0.004
					[0.016]	[0.016]
log(Sedentary minutes)	α_e			0.010	0.042	0.038
				[0.007]	[0.013]***	[0.013]***
G X Sedentary min.	$\alpha_{g\times e}$				0.026	0.023
					[0.016]*	[0.016]
$\log(BMI)_{t-1}$	$(1-\delta)$		0.990	0.947	0.947	0.972
			[0.011]***	[0.012]***	[0.012]***	[0.011]***
$\log(BMI)_{mom}$	γ_b		0.073	0.076	0.076	
			[0.011]***	[0.011]***	[0.011]***	
Covariates			X	X	X	
R^2		0.25%	78%	79%	79%	78%
Observations		3346	3346	3346	3346	3346

* significant at 10%; ** significant at 5%; *** significant at 1%. Standard error clustered at the individual level in brackets. Dependent variable: log BMI (kg/m^2); Covariates: parity; age of child at clinic date; mom and dad education and SES; mother age at pregnancy; dummy for single mother; reliable dietary report; time dummy; month dummies; late respondent; birth weight.

Table 15: Robustness Checks - Genetic Score

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Baseline	Males	Females	No Underweight	RE	FE	Prob Overweight	Weight	zBMI	Fat %
Risky Genetic Score	β_g	0.012	0.015	0.011	0.011	0.014		0.095	0.012	0.094	0.021
		[0.002]***	[0.004]***	[0.003]***	[0.003]***	[0.003]***		[0.063]	[0.003]***	[0.019]***	[0.018]
log(Food Int.)	α_f	0.065	0.076	0.072	0.063	0.057	0.012	0.544	0.063	0.472	-0.007
		[0.008]***	[0.011]***	[0.013]***	[0.008]***	[0.009]***	[0.010]	[0.211]***	[0.010]***	[0.060]***	[0.067]
G X Food Int.	$\alpha_{g\times f}$	0.025	0.022	0.035	0.025	0.024	-0.009	0.197	0.020	0.210	-0.043
		[0.011]**	[0.015]	[0.017]**	[0.011]**	[0.012]*	[0.014]	[0.269]	[0.013]	[0.079]***	[0.088]
log(Sedentary m.)	α_e	0.019	0.020	0.014	0.021	0.032	0.030	0.519	0.028	0.132	0.085
		[0.008]**	[0.012]*	[0.012]	[0.008]**	[0.009]***	[0.011]***	[0.196]***	[0.010]***	[0.060]**	[0.058]
G X Sedentary m.	$\alpha_{g\times e}$	0.000	-0.010	0.009	0.000	-0.006	-0.012	0.021	0.007	0.000	-0.077
		[0.011]	[0.015]	[0.015]	[0.011]	[0.011]	[0.014]	[0.242]	[0.012]	[0.077]	[0.078]
H_{t-1}	$(1-\delta)$	0.938	0.946	0.928	0.910	0.813	-0.136	2.102	0.759	0.868	0.306
		[0.008]***	[0.012]***	[0.011]***	[0.008]***	[0.009]***	[0.017]***	[0.053]***	[0.008]***	[0.008]***	[0.022]***
Controls		X	X	X	X	X	X	X	X	X	X
R^2		0.78	0.79	0.79	0.77		0.64		0.88	0.77	0.55
Observations		7,052	3,346	3,706	6,785	7,052	7,052	7,052	7,048	7,052	5,305

Column (1) reports the baseline estimates (same as table 6). Column (2) and (3) run the model separately for males and females. Column (4) runs the model dropping the children who are below the 5^{th} percentile of the z-BMI standard distribution for the UK (they represent 4% of the sample). Column (5) and (6) run the model using random effects and fixed effects, so that $\varepsilon_{i,t} = \mu_i + u_{i,t}$; all other columns report standard error clustered at the individual level. Column (7) runs a probit model on the probability of being obese. Column (8) uses H_t =log(weight) as dependent variable, controlling for log(height). Column (9) uses z-BMI as dependent variable. Column (10) uses the estimated percentage of body fat as dependent variable. For all the other columns, the dependent variable: log BMI (kg/m^2).

* significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors in brackets. Covariates: gender; parity; age of child at clinic date; mom and dad education and SES; mother age at pregnancy; dummy for single mother; reliable dietary report; time dummy; month dummies; late respondent; birth weight.

Table 16: Different Measures of Diet and Food Intake - FTO gene

		(1) Calories	(2) Proteins	(3) Fat	(4) Carbs	(5) Dietary Cholesterol	(6) Sugar	(7) Starch	(8) Non Starch	(9) Factor Score
Risky FTO	β_g	0.010	0.010	0.009	0.008	0.008	0.007	0.008	0.008	0.009
Gene		[0.002]***	[0.002]***	[0.002]***	[0.002]***	[0.002]***	[0.002]***	[0.002]***	[0.002]***	[0.002]***
log(Diet)	α_f	0.067	0.046	0.037	0.047	0.010	0.011	0.046	0.022	0.016
		[0.009]***	[0.007]***	[0.007]***	[0.008]***	[0.004]***	[0.005]**	[0.007]***	[0.005]***	[0.002]***
G X Diet	$\alpha_{g\times f}$	0.025	0.027	0.015	0.013	0.009	0.002	0.011	0.014	0.006
		[0.011]**	[0.009]***	[0.008]*	[0.010]	[0.005]*	[0.006]	[0.009]	[0.007]**	[0.003]**
log(Sedentary min.)	α_e	0.027	0.025	0.027	0.026	0.024	0.024	0.027	0.025	0.028
		[0.009]***	[0.009]***	[0.009]***	[0.009]***	[0.009]**	[0.009]**	[0.009]***	[0.009]***	[0.009]***
G X Sedentary min.	$\alpha_{g\times e}$	0.012	0.010	0.013	0.011	0.010	0.010	0.011	0.010	0.013
		[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]
$\log(BMI)_{t-1}$	$(1-\delta)$	0.939	0.939	0.944	0.942	0.945	0.946	0.943	0.947	0.937
		[0.008]***	[0.008]***	[0.008]***	[0.008]***	[0.008]***	[0.008]***	[0.008]***	[0.008]***	[0.008]***
Covariates		X	X	X	X	X	X	X	X	X
R^2		78%	78%	78%	78%	78%	78%	78%	78%	78%
Observations		7052	7052	7052	7052	7051	7052	7052	7052	7051

Column (1) reports the baseline estimates (same as table 3). The different measures of dietary intake used are: food intake (kilocalories/day - column 1); protein intake (grams/day - column 2); fat intake (grams/day - column 3); carbohydrate intake (grams/day - column 4); dietary cholesterol intake (mg/day - column 5); total sugar intake (grams/day - column 6); starch intake (grams/day - column 7); non-starch polysaccharide (fibre) intake (grams/day - column 8); factor score of all the dietary measures (column 9);

* significant at 10%; ** significant at 5%; *** significant at 1%. Standard error clustered at the individual level in brackets. Dependent variable: log BMI (kg/m^2); Covariates: gender; parity; age of child at clinic date; mom and dad education and SES; mother age at pregnancy; dummy for single mother; reliable dietary report; time dummy; month dummies; late respondent; birth weight.

Table 17: Different Measures of Physical Activity - FTO gene

		(1) Sedentary min	(2) MVPA	(3) Counts per min	(4) Factor Score
Risky FTO Gene	β_g	0.010	0.009	0.009	0.009
		[0.002]***	[0.002]***	[0.003]***	[0.002]***
log(Food Intake)	α_f	0.067	0.068	0.069	0.069
		[0.009]***	[0.009]***	[0.009]***	[0.009]***
G X Food Intake	$\alpha_{g\times f}$	0.025	0.021	0.024	0.023
		[0.011]**	[0.011]*	[0.011]**	[0.011]**
log(Exercise)	α_e	0.027	-0.011	-0.028	-0.008
		[0.009]***	[0.002]***	[0.005]***	[0.002]***
G X Exercise	$\alpha_{g\times e}$	0.012	-0.001	-0.009	-0.002
		[0.011]	[0.002]	[0.006]	[0.002]
$\log(BMI)_{t-1}$	$(1-\delta)$	0.939	0.934	0.936	0.936
		[0.008]***	[0.008]***	[0.008]***	[0.008]***
Covariates		X	X	X	X
R^2		0.78	0.79	0.79	0.79
Observations		7052	7043	7052	7043

Column (1) reports the baseline estimates (same as table 3). The different measures of exercise used are: sedentary minutes (column 1); moderate to vigorous physical activity (MVPA - column 2); counts per minute (column 3) factor score of all the exercise measures (column 4);

* significant at 10%; ** significant at 5%; *** significant at 1%. Standard error clustered at the individual level in brackets. Dependent variable: log BMI (kg/m^2); Covariates: gender; parity; age of child at clinic date; mom and dad education and SES; mother age at pregnancy; dummy for single mother; reliable dietary report; time dummy; month dummies; late respondent; birth weight.