Genetic and Economic Interaction in the Formation of Human Capital. The Case of Obesity

Pietro Biroli *

This Draft: October 2, 2013
PRELIMINARY DRAFT,
PLEASE DO NOT CIRCULATE

Abstract

I develop a model of health and human capital formation that takes into account the dynamic interaction between genetic inheritance and parental choices of investment in children. Differences in the genetic makeup of children can induce variations in the implicit cost of inputs and in the production function of human capital. In equilibrium this is mirrored by changes in the incentives to invest. I take the model to the data using the Avon Longitudinal Study of Parents and Children and I focus on a particular facet of health: obesity. Different forms of investment are considered as inputs, notably physical exercise and dietary intake, and I evaluate their interaction with specific genes which have been associated to increases in Body Mass Index in Genome-Wide Association Studies. I find that Gene-Environment interaction (GxE) plays a fundamental role in human capital formation: investments have a different effect on the accumulation of BMI depending on the genotype of the child. Children who are endowed with a particular genetic makeup are at higher risk of obesity when overeating, and yet they tend to display a higher caloric intake. These results are consistent with the findings in genetics and molecular biology showing that the FTO gene is associated with the hypothalamic regulation of food intake, and shed light on the interdependence between genes and economic choices regarding parental investment and human capital formation.

*Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637; email, biroli@uchicago.edu.

1 Introduction

> Nature makes the boy toward, nurture sees him forward.
> *Mulcaster (1582)*
>
> Genes load the gun. Lifestyle pulls the trigger.
> *Dr. Elliott Joslin*

Every day a parent chooses how to spend time, effort, and money in order to cultivate the natural predisposition of her child and give him a better chance in life. The goal of this paper is to understand how family choices of investment in the human capital of the child build on and interact with his genetic endowment in order to enable the full flourishing of his innate ability.

This work is nested within the debate on "nature versus nurture", which has long been discussed in social and biological sciences. It was initially framed as an antagonistic relationship by Sir Francis Galton, who believed that genetic inheritance played the stronger role: " When nature and nurture compete for supremacy on equal terms [...], the former proves to be the stronger." (Galton, 1874, p. 12) On the contrary in The Wealth of Nations Adam Smith argues that observed differences are due to specialization and division of labor, rather than arising from natural talents. "The difference between the most dissimilar characters, between a philosopher and a common street porter, for example, seems to arise not so much from nature, as from habit, custom, and education. When they come into the world, and for the first six or eight years of their existence, they were, perhaps, very much alike" (Smith, 1776, pp.28-29)

Since then a wide literature in behavioral genetics has tried to parse out the relative importance of these two components. Usually comparing identical and non-identical twins, these studies try to determine the precise percentage of a trait that is 'heritable'.[1] In a horse race between nature and nurture, they try to pick a winner. For example, it has been estimated that roughly 80% of the variation in human height can be attributed to genetic inheritance (Yang et al. (2010)), that 28 to 85% of IQ is heritable (van der Sluis et al. (2008)), or that 16-85% of Body Mass Index is due to genes (Yang et al. (2007)), while the rest is due to environment.

I argue that such an antagonistic relationship is ill-posed and obsolete. It should be relinquished in favor of a more systemic view that considers the dynamic interaction between the genes of an individual, and the environment in which he develops. This perspective conforms to the original idea of Richard Mulcaster, who first spoke of nature and nurture in more harmonious terms, stressing their collaborative effects. This

[1] The workhorse model decomposes the variance of a trait into three latent orthogonal components: Additive genetic, Common Environment, and unique Environment (ACE). The main idea is that, if mono-zygotic twins are more similar than fraternal twins in a particular domain, then this trait is highly 'heritable' and genetically determined. See (Plomin et al., 2008, ch. 5) for a textbook description. A critique of the main assumptions of this analysis has been provided by Goldberger (1976, 1979) and Manski (2011). A bivariate extension than joins research from social and biological sciences has been proposed in Kohler et al. (2011), who discuss at length the necessary identifying conditions.

view has recursively been put forth by other scholars[2], and it is deeply rooted in the evidence accumulated in genetics and molecular biology. Recent technological advances in the mapping of the genome allowed researchers to connect various human traits to specific genetic markers. Through appropriate statistical models and using Genome-Wide Association Studies (GWAS), these discoveries contributed to our understanding of the genetic underpinning of human behaviors and characteristics. At the same time the epigenetic work of Meany, Syzf and colleagues (Meaney and Szyf (2005); Meaney (2010); Szyf and Bick (2013)) has shown how the genetic endowment of an individual is shaped by the surrounding environment through the process of gene expression and DNA methylation, validating the claim that the environment 'gets under your skin.' These strands of research combined gave a biological foundation to the studies regarding Gene-Environment interaction (GxE) and Gene-Environment correlation (rGE).[3] The interplay between genes and the surrounding environment could explain the change in certain human traits that are genetically determined, but yet rose significantly in the past decades against a constant gene pool. Dickens and Flynn (2001) show how GxE combined with high levels of heritability and strong environmental gains can explain the increase in IQ witnessed in the past 40 years; Yeo and Heisler (2012) describe how shifts in lifestyle combined with genetic responses to obesogenic environment can account for the recent rise in obesity rates.

I contribute to the debate by introducing an economic framework of investment in the capabilities of the child, conditional on his genetic endowment. The family shapes the environment of the child in order to develop his full potential, while facing limits on the time and the resources available.

In order to achieve this goal, I develop a simple model of health and human capital formation, taking into account the dynamic interaction between genetic inheritance and family choices of investment. The cardinal idea is to model the genetic code of an individual as delimiting his possibility set. The DNA is a biological shifter of the subjective cost of investments. In other words, the genetic endowment of an individual delineates the set of achievable combination of inputs (investments) and outputs (human capital) that an agent can attain; therefore a genotype that has been related to a particular trait, such as cognitive ability[4], would entail a change in the production set faced by the individual, and better chances to attain a higher cognitive standing. The actual achievement of such ability, however, would depend on the sequence of economic choices and investments undertook in order to develop that particular trait.

[2]See for example Anastasi (1958), West and King (1987), and more recently Rutter (2006) and Heckman (2007)

[3]Gene-environment interaction is used to describe the genetic differences in the sensitivity or vulnerability to a particular environment. Gene-environment correlations refer to the genetic influences on the likelihood that individuals will experience certain environments. These terms were initially introduced and defined by Plomin et al. (1977). See Moffitt et al. (2005, 2006) and (Rutter, 2006, ch.9) for a through explanation of these terms.

[4]See Davies et al. (2011); Butcher et al. (2006) for a discussion of the genes that have been associated with IQ and cognitive functioning

This assumption puts some structure on how the genetic code of an individual interacts with the budget constraint faced by the family, and generates predictions that can be tested by the data. Considering various types of investments, I evaluate how differences in the genotype of the child induce variations in the implicit cost of inputs. In equilibrium this is mirrored by changes in the returns to investments and the optimal allocation of resources within the family.

The prediction of the model are tested using the Avon Longitudinal Study of Parents And Children (ALSPAC), a very rich epidemiological dataset that followed prospectively a birth cohort recruited in Avon, UK, in 1991/1992.

This paper focuses on a particular facet of health capital: obesity. A fundamental determinant of productivity, wellbeing, and longevity, child obesity in the US more than doubled in the last decades[5], rising so dramatically to be dubbed 'obesity epidemic'. Furthermore obesity is well suited for this analysis because it has strong genetic basis, yet it can be influenced by individual choices on diet and physical exercise. To guide the empirical analysis, precise measures of genes and environment are selected.[6] The choice of the gene is driven by findings in molecular biology, which show how minor variants in the FTO gene - rs9939609 Single Nucleotide Polymorphism (SNP) - are associated with hypothalamic regulation of energy intake but not energy expenditure.[7] In other terms, this particular gene has been linked to biological mechanisms in the brain that determine the control of appetite and feeding impulses, but it has not been associated with differences in the rate of calories burnt. Therefore the presence of minor-allele variants of this gene increases the cost of following a strict diet, without altering the incentives to engage in physical activity.

This intuition is corroborated by the results: there is strong evidence of the interaction between FTO and the quantity of calories ingested by the child; furthermore, the data shows an interaction between the FTO genotype and physical activity for males, albeit to a minor extent. As predicted by the model the gene is connected, on average, to higher levels of Body-Mass-Index (BMI). I show how this average effect can by explained by the interaction between FTO and the decisions of the family, changing both the implicit cost and the productivity of different investments. These results are replicated when considering a genetic score of predisposition to obesity based on 26 different genes. The presence of a different genotype does not predestine the child to be overweight, but rather his level of health is conditional on the family choices in the

[5]See Ogden et al. (2002)

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

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

realm of diet and exercise.

2 The Model

I consider a simple model of health and human capital formation where the family has to decide how many resources to invest in the development of the child.[8] Building on the work of Cunha and Heckman (2007) and Conti and Heckman (2010), I analyze a simple model to derive the basic predictions and build some intuition.

Consider the family acting as a firm: they produce human capital H, using investments I_k, subject to a budget constraint. They face the following problem:

$$\begin{aligned} \max_{I_d, I_e} H_t &= f(I_d, I_e; g) + (1-\delta)H_{t-1} \\ s.t\ Y &= p_d(g)I_d + p_e(g)I_e \end{aligned} \tag{1}$$

H_t is the current stock of human capital, which is function of investments I_d and I_e, as well as the past stock H_{t-1}. I assume that the function $f(.)$ is strictly increasing and concave, and twice continuously differentiable in I_k.[9] For the case of obesity analyzed in the empirical section, consider H_t as the level of fitness of the child, and the two inputs as investment in diet I_d and in physical exercise I_e. The parameters of the production function $f(.)$ are indexed by the child's genotype g. Y are the family resources dedicated to the investment in the child's human capital; they are split between I_d and I_e according to their relative costs $p_d(g)$ and $p_e(g)$, which depend on the genetic endowment of the child. I assume that the family knows the relevant parameters of the production function and observes the cost of the investments; however, they are not necessarily aware of the genotype of the child. In other words, the family knows how costly it is to achieve a certain level of investment, and how useful such investments are in the formation of human capital; however, they do not need to know which particular alleles the child is endowed with.

A fundamental feature of the model is that genes affect the accumulation of human capital in two ways. First, they influence the production function $f(.;g)$: genes can change the way inputs are converted into outputs; for example g can shift the

[8] I am not taking a stand on who or how many are the decision makers in the household, and I do not try to model the internal decision making process of a family. Dauphin et al. (2011) show evidence that children aged 16 and older are also decision makers; comparing unitary and collective decision models for the household, they always fail to reject the collective model. Lundberg et al. (2009) estimate a non-cooperative model to study the decision-making by children distinguishing between decisions taken on their own and shared with their parents. They find that the probability of taking independent decisions increases sharply between age 10 and 14. Finally, Cosconati (2009) and Cosconati (2012) estimate a dynamic model of parent-child interaction that evaluate the effectiveness of different parenting styles in the formation of human capital. In this paper I abstract from this and I simply assume that the decision-making process of the family, whatever its exact nature, does not depend on the genetic endowment of the child and leads to Pareto efficient outcomes in terms of investment decisions.

[9] These conditions are sufficient for an internal optimum.

effectiveness of one input, or the substitutability between different investments. This can be called the genetic *productivity effect*. This effect is related to the literature on Gene-Environment interaction (GxE).[10] For example, consider a gene that facilitates the creation of neural connections: for the same level of investment in schooling and lectures, this gene will increase the cognitive ability of the child.[11] A second channel is the effect of genes on the cost of investments $p_k(g)$; 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. This can be labeled the genetic *cost effect*. This effect can be loosely related to Gene-Environment correlation (rGE).[12] For example, consider a gene that increases the attention and the focus of the child: this will make it easier to teach him new concepts, effectively reducing the cost of such investment.

Drawing a connection with the household production model developed by Becker, the genetic endowment would be considered as an 'environmental variable' which influences the household production function. As explained in (Becker, 2007a, p.48), such variables "reduce the cost of producing commodities, and thus would expand opportunities, even if the full income were not affected."

Indeed notice how both genetic effects are associated with an increase in the welfare of the family: 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.

The first order conditions (FOC) of the model relate the marginal rate of substitution between the investments to the ratio of costs:

$$\frac{\frac{\partial f(I_d,I_e;g)}{\partial I_e}}{\frac{\partial f(I_d,I_e;g)}{\partial I_d}} = \frac{p_e(g)}{p_d(g)} \tag{2}$$

Equation (2) shows the two main effects of genes: on costs in the right-hand-side,

[10]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 $\frac{\partial f}{\partial I_k}$ is a function of g

[11]A more famous example comes from the research of Caspi et al. (2002), who show how the gene MAOA moderates 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.

[12]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 cost $p_k(g)$ is lower for certain genotypes, we would expect them to be exposed to higher levels of I_k.

and on productivity in the left-hand-side. However these two effects could operate in opposite directions. In order to obtain clear implications on the equilibrium level of investment, the cost effect and the productivity effect of the same gene should not perfectly offset each other. A sufficient condition to avoid such indeterminacy is that a particular genotype does not influence the costs, or does not interact with the investment in the production function of human capital $f(.)$. This occurs if the ratio of costs p_e/p_d does not depend on g, or if the ratio of partial derivatives $f'_{I_e}(I_d, I_e; g)/f'_{I_d}(I_d, I_e; g)$ is independent of the child's genotype. For example this would occur if the cross partials evaluated at a different allele, $g = A$ or $g = T$, are equal to each other: $f'_{I_k}(I_d, I_e; g = A) = f'_{I_k}(I_d, I_e; g = T)$, for $k = e, d$.[13]

A more general condition can be derived by taking the difference of equation (2) evaluated at two different g. The productivity effect is $\frac{f'_{I_e}(I_d,I_e;g=A)}{f'_{I_d}(I_d,I_e;g=A)} - \frac{f'_{I_e}(I_d,I_e;g=T)}{f'_{I_d}(I_d,I_e;g=T)}$, while the cost effect is $\frac{p_e(g=A)}{p_d(g=A)} - \frac{p_e(g=T)}{p_d(g=T)}$.[14] The productivity effect does not overturn the cost effect if they have the same sign, or if one is smaller than the other in absolute value:

$$\left|\frac{f'_{I_e}(I_d, I_e; g = A)}{f'_{I_d}(I_d, I_e; g = A)} - \frac{f'_{I_e}(I_d, I_e; g = T)}{f'_{I_d}(I_d, I_e; g = T)}\right| < \left|\frac{p_e(g = A)}{p_d(g = A)} - \frac{p_e(g = T)}{p_d(g = T)}\right|$$

Everything simplifies when assuming a particular functional form for the production function $f(.)$. For example, consider a Cobb-Douglas: $f(I_d, I_e; g) = A(g)\left[I_e^{\alpha_e(g)} I_d^{\alpha_d(g)}\right]$. In this case α_k is a measure of the productivity of investment I_k, since it represents the percentage change in output divided by the percentage change in input (output elasticity). As derived in the appendix (A.1), the optimal level of investment is:

$$I_d^* = \frac{\alpha_d(g)}{p_d(g)} \frac{1}{\alpha_e(g) + \alpha_d(g)} Y \tag{3}$$
$$I_e^* = \frac{\alpha_e(g)}{p_e(g)} \frac{1}{\alpha_e(g) + \alpha_d(g)} Y$$

We can see that I_k^* is inversely related to $p_k(g)$, and directly proportional to $\alpha_k(g)$. Therefore investment k will increase if costs decrease with the gene, $p_k(g = A) < p_k(g = t)$, or if productivity increases with the gene, $\frac{\alpha_k(g=A)}{\alpha_j(g=A)} > \frac{\alpha_k(g=T)}{\alpha_j(g=T)}$. This is indeed quite

[13] Considering g as a continuous indicator of genetic predisposition, rather than a particular allele, we can take derivative with respect to g. The condition would then be the following: $\frac{\partial f(I_k,I_e;g=A)}{\partial I_e \partial g} = \frac{\partial f(I_d,I_e;g)}{\partial I_d \partial g}$. Notice that if g enters multiplicatively in the functions, the cross partials do not depend on the genotype. For example consider the case of a Hicks-neutral technical change: $H = f(I_d, I_e; g) = gf(I_d, I_e)$.

[14] When considering a continuous value of genes g, we can take the double derivative with respect to the investment and the genotype and compare the cross-partials. The productivity effect is $\frac{\partial f(I_d,I_e;g)}{\partial I_e \partial g}\frac{\partial f(I_d,I_e;g)}{\partial I_d} - \frac{\partial f(I_d,I_e;g)}{\partial I_d \partial g}\frac{\partial f(I_d,I_e;g)}{\partial I_e} / \left[\frac{\partial f(I_d,I_e;g)}{\partial I_d}\right]^2$ and the cost effect: $\frac{\partial p_e(g)}{\partial g} p_e(g) - \frac{\partial p_d(g)}{\partial g} p_d(g)/(p_d(g))^2$

intuitive: if an input is cheaper, or more productive, the family will demand more of it. A similar results holds when assuming a Constant-Elasticity-of-Substitution production, as shown in appendix (A.1.1).

A more complex model of household production, in the spirit of Grossman (1972), Grossman (2000), and Becker (2007a), can be used to better understand the source of what I call the cost effect. Consider a family that cares about consumption, leisure, and the human capital of the child. Human capital H is a function of different family investments I_k, which are produced inside the household using goods purchased in the market x_k as well as time and effort τ_k. The costs of these goods and the value of time do not depend on the genotype of the child. However the child's gene can influence the amount of time or the level of goods needed in order to achieve a certain level of investment, so that $I_k = I_k(x_k, \tau_k; g)$. In equilibrium, the implicit cost of I_k will be the sum of the money spent on purchasing goods x_k plus the value of time spent investing, $p'_k = p_x x^*_k + w\tau^*_k$. If the genotype of the child influences the optimal amount of time and effort needed to achieve a certain level of investment, so that $\tau^*_k(g)$ or $x^*_k(g)$ vary with g, then we would observe that the implicit cost of investment $p'_k(g)$ depends on genes as well.

Consider the example of cost effect provided above: a certain gene increases the attention and the focus of the child. Such increase in attention will require less time and effort to to teach him new concepts, so that $\tau^*_k(g)$ decreases with g, effectively reducing the implicit cost of investment k.

The First Order Conditions of this model are throughly derived in appendix (A.2). When focusing on the optimal level of investment, the following is obtained:

$$\frac{\frac{\partial f(I_d,I_e;g)}{\partial I_e(x_e,\tau_e;g)}}{\frac{\partial f(I_d,I_e;g)}{\partial I_d(x_d,\tau_d;g)}} = \frac{p_{x_e} / \frac{\partial I_e(x_e,\tau_e;g)}{\partial x_e}}{p_{x_d} / \frac{\partial I_d(x_d,\tau_d;g)}{\partial x_d}}$$

where $f'_{I_k} = \frac{\partial f(I_d,I_e;g)}{\partial I_k(x_k,\tau_k;g)}$ is the marginal productivity of investment k, while $p'_k = p_k / \frac{\partial I_k(x_k,\tau_k;g)}{\partial x_k}$ represents its implicit cost. This is very similar to the First Order Conditions of the simplified model presented before, and gives stronger foundations to equation (2).

A concrete example can help pin down the ideas and understand better the different parts of the model. Let's look at the gene that will be used in the empirical section of this paper: FTO. Various studies in molecular and human genetics have shown how FTO is associated to obesity through the regulation of appetite and hunger (energy intake), but it does not seem to influence metabolism and physical exercise (energy expenditure).[15] For example Fredriksson et al. (2008) use mice-models to analyze the

[15]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.

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.[16] Similar results are reported by Olszewski et al. (2009) and Tung et al. (2010), who find significant changes in the activity level of *Fto* in the hypothalamus of rats and mice experimentally deprived of food.[17] 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 gene variant] was associated with increased energy intake independently of body weight"; however it had no visible effect on their resting energy expenditure and metabolism. Analyzing the same dataset used in this paper, Timpson et al. (2008) find a strong effect of the A-allele on increased total energy 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 increases the effort needed to follow a strict diet, so that $I_d(A_{FTO}) \leq I_d(T_{FTO})$, but has no clear effect on the exercise function, so that $I_e(A_{FTO}) \approx I_e(T_{FTO})$. Therefore carrying at least one A allele increases the overall cost of investing in a diet, $p_d(A_{FTO}) \geq p_d(T_{FTO})$, but does not change the cost of exercise, $p_e(A_{FTO}) \approx p_e(T_{FTO})$: there is a genetic cost effect.[18] The increase in the cost of one investment will induce a shift in the budget set and, consequently, a change in the optimal allocation of both diet and exercise, as shown in Figure 1. 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 diet and, consequently, to a lower level of health (higher BMI). The consequences on activity are not straightforward but depend on the substitution between the two investments.

Regarding the evidence of a genetic productivity effect, Kilpeläinen et al. (2011) review 45 studies of adults and nine studies of children and adolescents that focus on

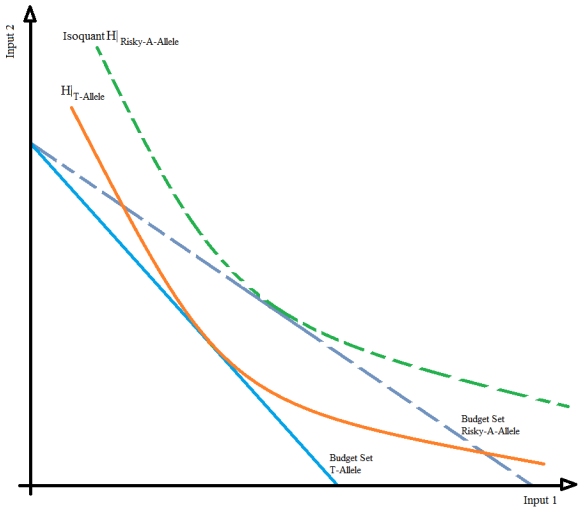
(2008) show results in different human populations.

[16]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."

[17]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.

[18]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 $\approx$. 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."

Figure 1: Genetic Cost Effect



the interaction between FTO and physical activity. They find that exercise attenuates the negative effect of FTO on obesity, especially in adults. In terms of the model, this means that the FTO gene increases the productivity of investing in physical exercise: $f'_{I_e}(g = A) \geq f'_{I_e}(g = T)$. For the same level of exercise, those with the FTO gene can attain a more favorable outcome (lower BMI), as shown in Figure (2).[19]

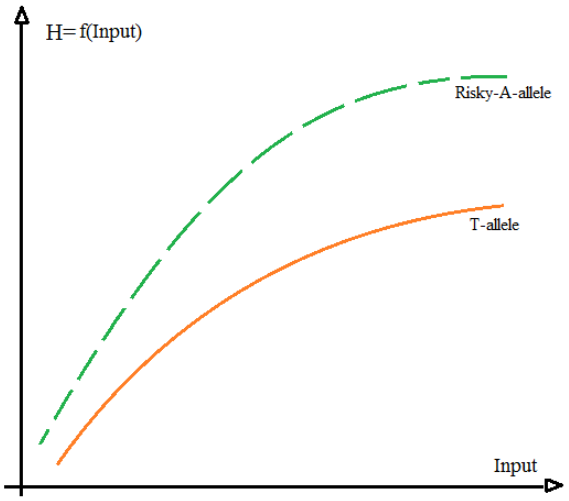
Although the case of FTO has been analyzed in full detail, very similar physiological functions have been found for other obesity-genes, such as MC4R, BDNF, SH2B1.[20]

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.

[19]It is worth noticing that the results provided in the meta-analysis of Kilpeläinen et al. (2011) do not control for diet and energy intake. In other words, they estimate the relation between H and I_e without considering the impact of I_d.

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

Figure 2: Genetic Productivity Effect



3 Empirical Results

The model can be applied 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 implications of the theory. The choice of these three components must be thoughtful and based on the existing evidence that connects investments and genotype to the final outcome of interest.[21] As mentioned before, I focus on obesity as measure of H, diet and exercise as proxies for I_k, and FTO and other genetic markers as assays of g.

Childhood health is an important and yet still underestimated facet of human capital.[22] Obesity in particular has become of prime importance due to the dramatic rise in fat 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.[23]A similar trend occurred in England: 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.[24]

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

[22]See Becker (2007b) and Heckman (2012)

[23]There is some evidence that obesity rates has stopped increasing in the last two years. More recent statistics can be found at `http://www.cdc.gov/obesity/data/childhood.html`

[24]It is worth noticing the level of obesity rates in the US and the UK reported in these studies are

The negative health consequences of obesity are far-reaching: child obesity contributes to the risk of developing metabolic syndrome, type 2 diabetes, hypertension, dyslipidaemia, non-alcoholic steatohepatitis, and obstructive sleep apnea;[25] furthermore it increases the chances of being overweight also during adulthood, with negative consequences on coronary heart disease, some types of cancer, and longevity.[26] Cawley (2010) shows how both direct and indirect costs of obesity are considerable: childhood obesity in the US costs $14.3 billion a year due to prescription drugs, emergency room, inpatient and outpatient costs; the figure for adults is 10 times greater, with an estimated $147 billion spent in obesity related illnesses.[27] The indirect costs of obesity range from delayed skills acquisition, to lower wages, job absenteeism, and lower productivity.

Finally, I chose obesity because it can be recorded easily, reliably, and consistently across space and over time, allowing intergenerational and cross-country comparisons; it has a strong biological underpinning that has been connected to various genetic markers; and yet the level of fitness of our body can be affected by both social and economic choices, such as diet and exercise.

The choice of the gene is driven by the molecular genetics findings mentioned before, the importance of FTO in determining obesity, and the wide prevalence of the risky allele. The intron 1 of the FTO gene was first related to susceptibility for obesity in genome-wide association studies (GWAS) by Frayling et al. (2007); the finding was then replicated by multiple authors using different datasets. Each additional minor allele of the rs9939609 single nucleotide polymorphism (SNP) in FTO 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.[28] The explanatory power of a single gene 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. The model suggests the same conclusion: a small change in the cost of an input will not induce a substantial change in the output. Indeed, 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

not comparable because they use a slightly different definition of obesity; Ogden and colleagues use a cutoff above the sex-specific 95^{th} percentile of the CDC's 2000 BMI-for-age growth chart; Stamatakis compute age- and sex-specific obesity rates based on the international definition of Cole et al. (2000), who compute percentile curves that passed through the cutoff point of 30 kg/m^2 at age 18, using data from 6 different countries.

[25] See among others Daniels et al. (2009)

[26] See Singh et al. (2008) for the tracking of childhood obesity into adulthood; for the negative consequence of adult obesity, read LeBlanc et al. (2011), Whitlock et al. (2009) and the references therein.

[27] Finkelstein et al. (2009) obtain this estimate 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.

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

close to 50% in genetically diverse populations. 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.[29] Using the words of Tung and Yeo (2011), "FTO could possibly be influencing the BMI of up to half the world's population!"

3.1 The Data

To bring the model to the data, I use the Avon Longitudinal Study of Parents and Children (ALSPAC), an ongoing investigation on the health and development of young children (`http://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 with an expected delivery date between April 1991 and December 1992 (Golding et al. (2001)). 14,541 pregnant women were enrolled at the beginning of the study.

Health and lifestyle data were collected through regular questionnaires, as well as medical and educational records. Anthropometric, Physical Activity, Dietary, and dual-energy X-ray absorptiometry (DXA) measures were obtained during research clinic visits.

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.[30] 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.[31]

Dietary assessment

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

[29]Estimates reported by Kilpeläinen et al. (2011) using HapMap population.

[30]See Cole et al. (1998)

[31]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.

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.

The summary statistics of the main variables can be found in tables (1-2). As expected, the A-allele of the FTO gene is related to higher levels of Body Mass Index for both males and females, and for higher levels of energy intake for males. Very similar results can be found when looking at different measures of adiposity, physical activity, and food intake.[32]

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

	Body Mass Index				Sedentary Hours			
	Female		Male		Female		Male	
Age	T-Allele	A- Allele	T-Allele	A- Allele	T-Allele	A- Allele	T-Allele	A- Allele
11	18.57	19.03	18.16	18.59	7.18	7.25	6.89	6.98
	(10.12)	(10.35)	(8.10)	(9.70)	(1.19)	(1.21)	(1.27)	(1.45)
13	20.42	20.87	19.74	20.08	8.26	8.24	7.73	7.77
	(11.86)	(12.57)	(10.29)	(11.68)	(1.32)	(1.31)	(1.50)	(1.54)

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

3.2 Evidence of interaction between genes and investment

First of all let us look at 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).

[32]See table (10) in the appendix.

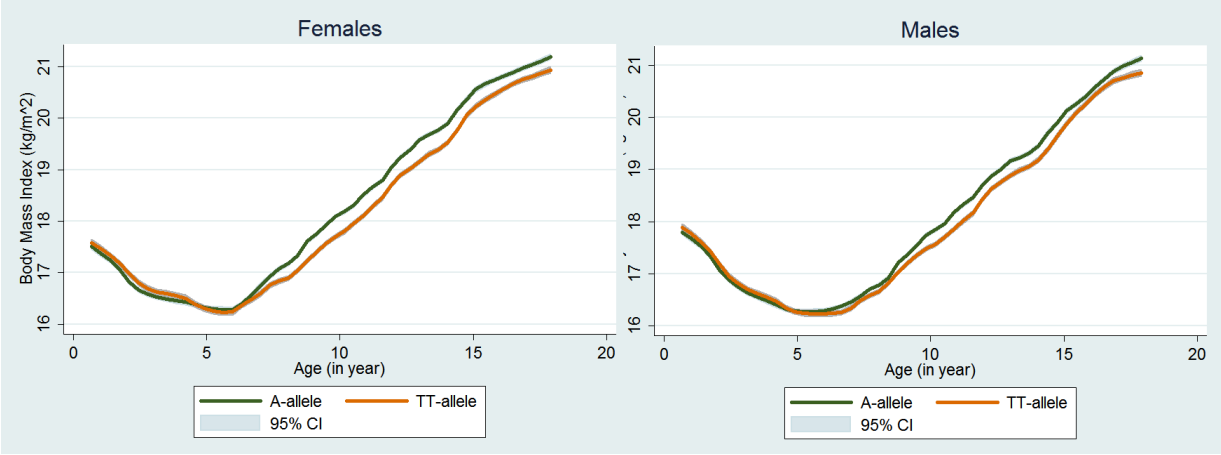
Table 2: Summary Statistics by age, gender, and genotype

	Kilocalories				Whole Sample		
	Female		Male				
Age	T-Allele	A- Allele	T-Allele	A- Allele	BMI	Sed	Kcal
11	1.75	1.78	1.92	1.97	18.65	7.10	1.86
	(0.13)	(0.12)	(0.15)	(0.16)	(9.78)	(1.31)	(0.15)
13	1.77	1.76	2.12	2.15	20.34	8.02	1.95
	(0.21)	(0.18)	(0.30)	(0.27)	(11.93)	(1.47)	(0.27)

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

Figure 3 depicts the evolution of body-mass-index from birth to age 20; 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-gene 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 gene becomes pronounced as the effect of environment accumulates over time.

Figure 3: Evolution of Body-Mass-Index



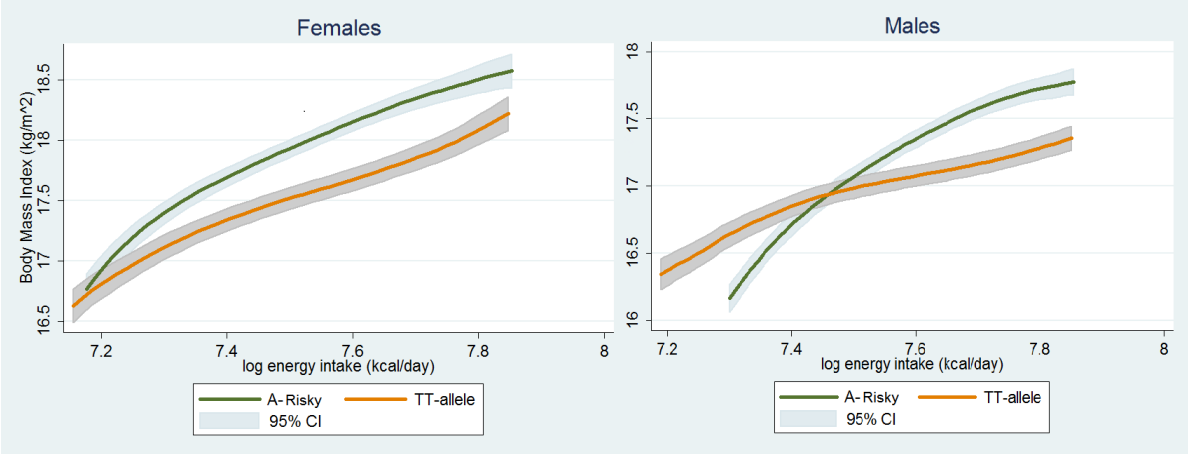
Nonparametric local-mean smoothing using Epanechnikov kernel and Silverman's Rule-of-Thumb bandwidth. Combining information from successive clinical visits, age 0 to 20; excluding outliers in the top and bottom 5% of the BMI distribution.

Let's turn to analyzing the genetic productivity effect: we want to see whether, for the same levels of inputs I_k, a different level of output H is obtained depending on the genotype of the child. This is related to the so-called Gene-Environment interaction

(GxE).[33] 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 I_d by gender and genetic endowment, I investigate the relationship between investment in diet and the obesity of children aged 10 to 14 years old.[34] Figure 4 shows the relation between the logarithm of the total amount of energy intake (kilocalories per day) and the BMI using a local smoothed average. Not surprisingly, a higher energy 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 seems that $f'_{I_d}(g = A) \geq f'_{I_d}(g = T)$. Furthermore, the two slopes intercept at low levels of energy 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 genes and environment are present and interact.

Figure 4: Gene-Diet Interaction: the effect of energy 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(energy intake).

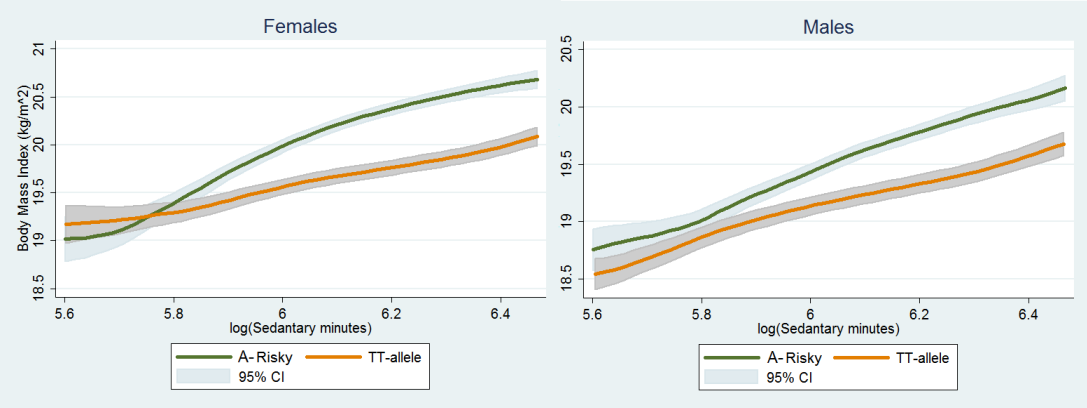
A similar result can be found by analyzing the level of physical activity and exercise chosen by the adolescents, as depicted in Figure 5. Again we see that the interaction between FTO and exercise is present, but it is not as pronounced as the interaction between FTO and diet. More time spent in a sedentary lifestyle leads to an increase in BMI, but significantly more so for the children who happen to carry the risky A-allele. That is, for those children who have a higher cost of diet, and therefore tend to eat more. This is consistent with the predictions of the model: FTO induces a higher cost of dieting; this will induce a higher level of energy intake; such intake will have

[33] See Moffitt et al. (2005, 2006) and in (Rutter, 2006, ch.9) for a through explanation of the term GxE.

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

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 5: Gene-Activity 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 diet, and FTO and exercise.[35] 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 inactive twins, 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.[36]

3.3 A Linear Production Function of Health

The analysis of the raw data gives suggestive evidence of gene-environment interaction, and the existence of a productivity effect. I now turn to a parametric estimation that allows 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 entire evolution of human capital.

First of all I consider a linear production function, log-linearizing a Cobb-Douglas function of the form: $H_t = A(X)g^{\beta_g}\left[I_{e,t}^{\alpha_d(g)} \cdot I_{d,t}^{\alpha_e(g)}\right]$. Following Ehrlich and Chuma

[35]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.

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

(1990) and Galama et al. (2012), I allow for decreasing returns to scale, so that the αs do not need to sum to 1. The theoretical model suggests that the child's genotype can modify the structural relation between investments and health. The genes can change the overall productivity of the inputs, entering multiplicatively in the production function: such change would be captured by a change in the constant β_g. This is equivalent to a Hicks-neutral technical change and would be represented by a vertical shift in figures (4) and (5). At the same time the genotype can change the marginal productivity of a particular investment: such change would be captured by different values of α_d and α_e. This is equivalent to a non-neutral technical change and would be represented by a shift in the slopes depicted in the figures above. Since the evidence of the previous section (3.2) is in favor of the latter, I introduce an interaction term between g and both I_d and I_e. The coefficients of these interactions between genes and the investment can be used to test the existence of a productivity effect.

Finally the term $A(X)$ controls for additional variables that can influence the child's BMI. In particular I consider the persistence of health capital by introducing $(1-\delta)H_{t-1}$, where δ is the depreciation rate of the health stock; and I test for intergeneration transmission of human capital by controlling for mother's genetic endowment and health characteristics. The following log-linear equation is estimated:

$$\begin{aligned} log(H_{i,t}) = \mu + \beta_g g + \alpha_e log(I^e_{i,t}) + \alpha_d log(I^d_{i,t}) + \\ + \alpha_{g\times e} log(I^e_{i,t}) \cdot g + \alpha_{g\times d} log(I^d_{i,t}) \cdot g + (1-\delta) log(H_{i,t-1}) \\ + \gamma_g g^{mom}_i + \gamma_h log(H^{mom}_i) + \beta X_{i,t} + \kappa_t + \varepsilon_{i,t} \end{aligned} \tag{4}$$

$H_{i,t}$ is the BMI of child i at time t; g is a dummy indicating the genotype of the child; $I^d_{i,t}$ are kilocalories consumed, as a measure of investment in diet; $I^e_{i,t}$ are sedentary minutes, as a measure of investment in exercise; the coefficients $\alpha_{g\times d}$ 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 gene; $g_{mom,i}$ and $BMI_{mom,i}$ represent the genetic endowment of the mother and her weight before pregnancy; X_{it} are control variables introduced to proxy for family and individual specific characteristics that might influence obesity and investment[37]; κ_t captures age effects.

To summarize, the α coefficients capture the relevant parameters of the production function of health; δ is the depreciation rate of the health stock; the γ parameters capture the level of intergenerational transmission of human capital; the β capture the influence of demographic controls.

Since males and females have different metabolisms, different levels of BMI, and different habits, the model is estimated first pooled, and then separately for boys and girls in order to capture potential gender differences in the production function of health. As a measure of genetic endowment, 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

[37]I control for the age of the child 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

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)[38] 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).[39] Mendel's law of independent assortment states that different genes are uncorrelated, and indeed I find that all of these genetic loci have a very small correlation.[40] Consequently the genetic score displays a bell-shape similar to a normal distribution, as shown in figure 6. The underlying assumption in using this score is that each allele has the same marginal effect. In order to relax this assumption, 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 6: Distribution of the Genetic-Predisposition-Score

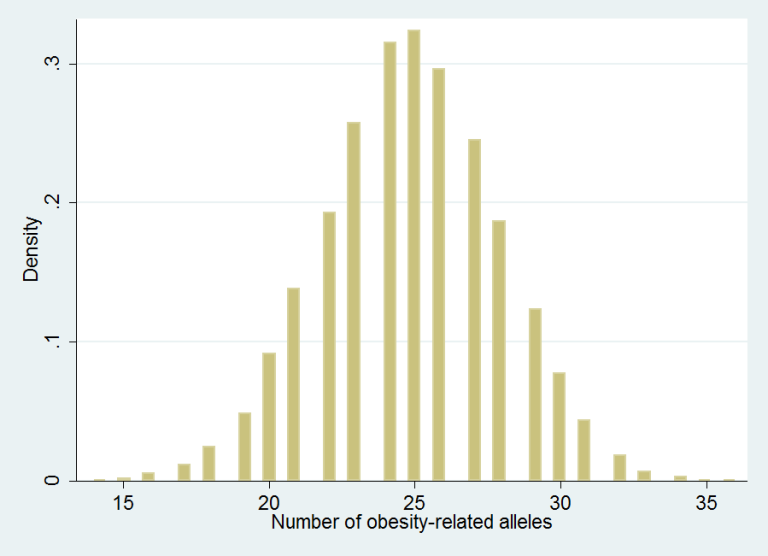


Table (3) reports the coefficients of equation (4) when using the risky A-Allele of the FTO gene as index of g.

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

[39]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 energy intake (diet) or energy expenditure (exercise). See the discussion in section 2.

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

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 and her genotype. The effect is similar to the ones found by related studies[41], 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 gene. Controlling for the investment choices of the family in column (3), the coefficient β_g does not change. Once the interaction between the gene and the two types of investment is introduced, the direct effect of FTO, β_g, still does not change and is equal to about a 1% change in BMI.[42] The most important contribution to BMI comes from diet and exercise, and their interaction with the genetic endowment. In this respect an interesting difference between males and females emerges, as shown in columns (2) and (3) of table (4); for girls the most important investment is the dietary decision, which also display a significant interaction with the FTO gene; their sedentary behavior is not predictive of BMI[43]. This is not true for boys, for whom both diet and exercise play and important role, and only the latter seem to interact with their genetic endowment. Notably, these effects are still very similar even when omitting the controls X in column (5).

Finally it is worth noting the very high persistence of BMI: the depreciation coefficient δ is very low, slightly more so for boys that girls. In other words, the past stock of human capital is a very important determinant of the current level of H_t.

Besides running the model separately by gender, various robustness checks are reported in table (4). So far I have considered BMI as a uniformly negative measure of health capital: the lower, the better. However this might not be a feasible characterization for those children with a very low level of body-mass. Less than 4% 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 mu_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 H_t as dependent variable; column (7) estimates of the probability of being overweight[44]; column (8) estimates the effect of genes and investment on changes in weight, control-

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

[42] The genotype g is interacted with the de-meaned investment variables, $I_k - \bar{I_k}$, so that the coefficient β_g captures the effect at the average of the investment distribution.

[43] It's worth noticing that once I control only for sedentary minutes and not for kilo-calories, then the investment in exercise becomes significant also for girls.

[44] 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.

Table 3: Gene and Investment Interaction - FTO

		(1)	(2)	(3)	(4)	(5)
Risky FTO Gene	β_g	0.019	0.006	0.006	0.009	0.010
		[0.005]***	[0.002]***	[0.002]***	[0.003]***	[0.003]***
log(Energy Intake)	α_d			0.052	0.068	0.069
				[0.006]***	[0.009]***	[0.009]***
G X Energy Intake	$\alpha_{g\times d}$				0.024	0.026
					[0.011]**	[0.011]**
log(Sedentary min.)	α_e			0.010	0.029	0.024
				[0.007]	[0.009]***	[0.009]***
G X Sedentary min.	$\alpha_{g\times e}$				0.012	0.012
					[0.011]	[0.011]
Mom Risky Gene	γ_g		-0.003	-0.003	-0.003	
			[0.003]	[0.003]	[0.003]	
log(BMI_{mom})	γ_h		0.090	0.090	0.090	
			[0.008]***	[0.007]***	[0.007]***	
$log(BMI)_{t-1}$	$(1-\delta)$		0.971	0.940	0.940	0.967
			[0.007]***	[0.008]***	[0.008]***	[0.008]***
Controls			X	X	X	
R^2		0.32%	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 FTO gene $g = 1$ if rs9939609 gene variant contains one or more A-alleles; $g = 0$ otherwise; Controls: 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; late respondent; birth weight.

ling 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 gene has an influence on the whole distribution of BMI, not only the upper tail, as shown in figures (8-7) in the appendix.[45]

[45]Further robustness checks are shown in the appendix. Tables (12) and (13) report the estimation of the linear health production separately by gender. Table (14) reports the estimates for the robustness checks, using the genetic score as a measure of g. Tables (15) and (16) report the estimation of the linear health production function using different measures of investments, both for diet I_d and exercise I_e. The estimated productivity effects $\alpha_{g\times k}$ do not change also when introducing an interaction between the genotype of the child and the lagged level of BMI $H_{i,t-1}$, or an interaction between g and the mother's BMI. Also introducing quadric terms for investments I_d and I_e does not influence the empirical conclusions. All results available upon request.

Table 4: Robustness Checks - FTO

		(1) Baseline	(2) Males	(3) Females	(4) No Underweight	(5) RE	(6) FE	(7) Prob Overweight	(8) Weight	(9) zBMI	(10) Fat %
Risky FTO Gene	β_g	0.009	0.006	0.013	0.012	0.013		0.218	0.014	0.092	-0.011
		[0.003]***	[0.005]	[0.003]***	[0.003]***	[0.004]***		[0.068]***	[0.003]***	[0.020]***	[0.020]
log(Energy Int.)	α_d	0.068	0.067	0.084	0.069	0.060	0.015	0.497	0.071	0.491	0.033
		[0.009]***	[0.013]***	[0.014]***	[0.010]***	[0.010]***	[0.012]	[0.224]**	[0.012]***	[0.070]***	[0.078]
G X Energy Int.	$\alpha_{g\times d}$	0.024	0.003	0.045	0.029	0.022	-0.004	0.082	0.028	0.191	0.025
		[0.011]**	[0.016]	[0.018]**	[0.011]**	[0.013]*	[0.015]	[0.274]	[0.014]**	[0.083]**	[0.094]
log(Sedentary m.)	α_e	0.029	0.042	0.012	0.030	0.042	0.026	0.574	0.032	0.206	0.096
		[0.009]***	[0.013]***	[0.013]	[0.009]***	[0.010]***	[0.012]**	[0.217]***	[0.010]***	[0.066]***	[0.067]
G X Sedentary m.	$\alpha_{g\times e}$	0.012	0.027	-0.007	0.009	0.008	-0.016	0.063	0.009	0.075	0.022
		[0.011]	[0.016]*	[0.016]	[0.011]	[0.012]	[0.014]	[0.253]	[0.013]	[0.080]	[0.082]
H_{t-1}	$(1-\delta)$	0.940	0.947	0.929	0.912	0.815	-0.136	2.096	0.761	0.870	0.310
		[0.008]***	[0.012]***	[0.011]***	[0.008]***	[0.009]***	[0.017]***	[0.052]***	[0.008]***	[0.008]***	[0.023]***
Controls		X	X	X	X	X	X	X	X	X	X
R^2		0.78	0.79	0.78	0.77		0.64		0.88	0.77	0.54
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 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. 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. Risky FTO gene $g = 1$ if rs9939609 gene variant contains one or more A-alleles; $g = 0$ otherwise. Controls: 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; late respondent; birth weight.

Table (5) 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 $\alpha_{g\times d}$ 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.[46] Finally, the magnitude of the estimated persistence of BMI, δ, is comparable to the estimates in table (3).

3.4 A CES Production Function of Health

One key feature of the model is that the marginal rate of substitution between the two investments might depends on the genetic makeup of the child (see equation (2)). In order to test this prediction I estimate a CES-production function of health, allowing all of the parameters to differ across genders and across genetic endowment. I consider the following CES specification $H_t = A(X)g^{\beta_g}\left[\alpha(g)I_{d,t}^{\eta(g)} + (1-\alpha(g))I_{e,t}^{\eta(g)}\right]^{\nu/\eta(g)}$.

As before, I allow the multiplying constant $A(X)$ to depend on various demographic controls X, the depreciation rate of capital $(1-\delta)H_{t-1}$, and mother characteristics. Taking the logarithm we obtain the following equation to estimate:

$$\begin{aligned}\log H_{i,t} =& \beta_g g + \frac{\nu}{\eta(g)}\log\left[\alpha(g){I_{i,t}^d}^{\eta(g)} + (1-\alpha(g)){I_{i,t}^e}^{\eta(g)}\right] + \\ &+ (1-\delta)\log H_{i,t-1} + \gamma_h log(H_i^m) + \beta X_{i,t} + \kappa_t + \varepsilon_{i,t} \qquad (5)\end{aligned}$$

where the parameters $\alpha(g)$ and $\eta(g)$ are allowed to vary according to the child's genotype g; ν captures the returns to scale, so that $\nu < 1$ is evidence of decreasing returns; and the elasticity of substitution is captured by the parameter $\sigma(g) = \frac{1}{1-\eta(g)}$.

Assuming that the error term $\varepsilon_{i,t}$ follows a normal distribution, equation (5) can be estimated using maximum likelihood.[47] The sample is split according to gender and two different indicators of genetic endowment: whether the child carries at least one risky A-allele in the FTO gene, and whether the child has a genetic score higher than 25. The results are displayed in tables (6) and (7) respectively.

The elasticity of substitution is always bigger than one, although noisily estimated, implying that the two investments are substitutes. More interestingly, the elasticity of substitution is slightly higher for the children with a 'risky' genotype: as pointed out in equation 2, the genetic endowment of the child can influence the elasticity of substitution between the two inputs. This is true for all the specifications, both pooled and divided by gender. However this difference is not statistically significant. Furthermore there is strong evidence in support of decreasing returns to scale (ν is always smaller

[46] For the robustness checks, see table (14) in the appendix.

[47] See section (B) for the details of the likelihood function.

Table 5: Gene and Investment Interaction - Genetic Score

		(1)	(2)	(3)	(4)	(5)
Risky Genetic	β_g	0.034	0.010	0.010	0.013	0.012
Score		[0.005]***	[0.002]***	[0.002]***	[0.003]***	[0.002]***
log(Energy Intake)	α_d			0.051	0.065	0.066
				[0.006]***	[0.008]***	[0.008]***
G X Energy Intake	$\alpha_{g\times d}$				0.025	0.026
					[0.011]**	[0.011]**
log(Sedentary min.)	α_e			0.010	0.021	0.014
				[0.007]	[0.008]**	[0.008]*
G X Sedentary min.	$\alpha_{g\times e}$				0.000	-0.003
					[0.011]	[0.011]
Mom Risky Gene	γ_g		-0.003	-0.003	-0.002	
			[0.002]*	[0.002]	[0.002]	
$\log(BMI_{mom})$	γ_h		0.090	0.090	0.090	
			[0.007]***	[0.007]***	[0.007]***	
$log(BMI)_{t-1}$	$(1-\delta)$		0.968	0.939	0.939	0.965
			[0.007]***	[0.008]***	[0.008]***	[0.008]***
Controls			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; Controls: 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; late respondent; birth weight.

Table 6: CES-Production function, by gender and FTO allele

		(1) All	(2) Females	(3) Males
Risky FTO	β_g	0.007	0.015	0.004
Gene		(0.005)	(0.004)	(0.001)
log(Energy Int.)	α_d	0.207	0.643	0.356
		(0.055)	(0.092)	(0.270)
G x Energy Int.	$\alpha_{d\times g}$	0.098	0.104	0.144
		(0.072)	(0.113)	(0.222)
Elasticity	η	0.390	0.339	0.232
of		(2.980)	(3.123)	(6.474)
substitution	η_g	0.050	0.016	0.129
		(2.634)	(3.774)	(5.015)
Returns	ν	0.298	0.464	0.450
to scale		(0.018)	(0.017)	(0.016)
$\log(BMI_{t-1})$	$(1-\delta)$	0.944	0.943	0.966
		(0.001)	(0.004)	(0.017)
$\log(BMI_{mom})$	γ_h	0.095	0.088	0.066
		(0.009)	(0.010)	(0.003)
Mom Risky	γ_g	-0.017	-0.009	-0.001
Gene		(0.002)	(0.002)	(0.000)
Elasticity of	$1/(1-\eta)$	*1.638*	*1.514*	*1.303*
substitution	$1/(1-\eta_g)$	*1.786*	*1.552*	*1.566*
Controls		X	X	X
Observations		7052	3706	3346
Likelihood		-8084.9	-4186.1	-3957.8

Dependent variable: log BMI (kg/m^2). MLE estimation, pooled and separated by gender. $\alpha_g = \alpha(g = Risky) - \alpha(g = Low - Risk)$ is the genetic difference in the parameter α. $\eta_g = \eta(g = Risky) - \eta(g = Low - Risk)$ is the genetic difference in the parameter η. Risky FTO gene $g = 1$ if rs9939609 gene variant contains one or more A-alleles; $g = 0$ otherwise. Controls: mom and dad SES; birth weight; age of child at clinic date; reliable dietary report; time dummy. Standard errors in parenthesis calculated as the inner product of the Hessian and the gradient of the likelihood.

Table 7: CES-Production function, by gender and Genetic Score

		(1) All	(2) Females	(3) Males
Risky Genetic	β_g	0.012	0.008	0.010
Score		(0.002)	(0.004)	(0.005)
log(Energy Int.)	α_d	0.467	0.622	0.527
		(0.043)	(0.125)	(0.112)
G x Energy Int.	$\alpha_{d\times g}$	0.078	0.081	0.156
		(0.003)	(0.000)	(0.029)
Elasticity	η	0.412	0.122	0.195
of		(1.530)	(4.815)	(2.750)
substitution	η_g	0.173	0.163	-0.021
		(0.470)	(1.182)	(0.194)
Returns	ν	0.378	0.719	0.438
to scale		(0.076)	(0.231)	(0.048)
$\log(BMI_{t-1})$	$(1-\delta)$	0.950	0.924	0.960
		(0.000)	(0.002)	(0.006)
$\log(BMI_{mom})$	γ_h	0.076	0.112	0.061
		(0.008)	(0.011)	(0.008)
Mom Risky	γ_g	-0.015	0.000	-0.011
Gene		(0.003)	(0.000)	(0.006)
Elasticity of	$1/(1-\eta)$	*1.701*	*1.138*	*1.243*
substitution	$1/(1-\eta_g)$	*2.412*	*1.399*	*1.212*
Controls		X	X	X
Observations		7052	3706	3346
Likelihood		-8113.6	-4156.3	-3962.4

Dependent variable: log BMI (kg/m^2). MLE estimation, pooled and separated by gender. $\alpha_g = \alpha(g = Risky) - \alpha(g = Low - Risk)$ is the genetic difference in the parameter α. $\eta_g = \eta(g = Risky) - \eta(g = Low - Risk)$ is the genetic difference in the parameter η. Risky genetic score $g = 1$ if genetic score > 25; $g = 0$ otherwise. Controls: mom and dad SES; birth weight; age of child at clinic date; reliable dietary report; time dummy. Standard errors in parenthesis calculated as the inner product of the Hessian and the gradient of the likelihood.

than 1). This was the case also in the Cobb-Douglas estimation, since $\alpha_d + \alpha_e < 1$. Finally the direct productivity effect of the genes, captured by the constant β_g, is similar to the one estimated before in table (3). There is strong evidence supporting the claim that the genotype changes the underlying characteristics of the production function, influencing the relative importance of the two investments.

It is worth noticing that there are seizable differences between genders: the share α of investment in diet (kilo-calories consumed) is slightly higher for females; finally, the intergenerational transmission of BMI, depicted by γ_h, is slightly higher for females.

3.5 Genetic Effects on the Investments

The theoretical analysis and the model suggest 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_k^* chosen by the family. As shown in equation (3), the optimal level of investment depend on the resources and the characteristics of the family, as well as from the genetic productivity effect $\alpha(g)$ and the genetic cost effect $p(g)$.

In order to test these predictions, I estimate equation (3) for both diet and exercise. The results of a regression of investments I_k on the genotype g of the child, lagged level of BMI $H_{i,t-1}$, and family resources and characteristics X are shown in table (8).

Table 8: Genetic Effect on Investments - FTO

	Caloric Consumption		**Sedentary Minutes**	
	Male	Female	Male	Female
	(1)	(2)	(3)	(4)
Risky FTO Gene	0.023	0.016	0.007	0.006
	[0.010]**	[0.009]*	[0.007]	[0.006]
Mom Risky Gene	0.016	-0.009	0.008	0.008
	[0.013]	[0.012]	[0.011]	[0.008]
$log(BMI_{t-1})$	-0.106	-0.160	0.077	0.052
	[0.036]***	[0.031]***	[0.025]***	[0.021]**
Controls	X	X	X	X
Observations	3,346	3,706	3,346	3,706

* significant at 10%; ** significant at 5%; *** significant at 1%. Standard error clustered at the individual level in brackets. Dependent variable: logarithm of daily kilocalories intake and daily sedentary minutes; Controls: 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; late respondent; birth weight.

As we can see, there is a significant genetic difference in the level of energy 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. Indeed, the previous results on the estimation

of the production function of health suggest that there is a genetic productivity effect of caloric intake: since $\hat{\alpha}_{g\times d} \geq 0$, eating more induces a higher level of BMI *especially* for those children who carry at least one A-allele of the FTO gene. This puts them at greater risk of obesity, and should decrease their level of energy intake. Indeed, equation (3) shows how I_d^* and $\alpha_d(g)$ are positively related: a higher risk of obesity should induce a higher effort in dieting. On the other side, I_d^* is negatively related to $p_d(g)$. Since we observe that children with the FTO A-allele tend to eat more, in spite of their increased risk, this suggests that $p_d(g = A) \geq p_d(g = T)$. In other terms, 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 energy intake but not energy expenditure.

The results are somewhat different when considering the effect of the genetic score on investments. Table (9) 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 genes used to construct this genetic score do not work through the same biological pathways as FTO.

Table 9: Genetic Effect on Investments - Genetic Score

	Caloric Consumption		**Sedentary Minutes**	
	Male	Female	Male	Female
	(1)	(2)	(3)	(4)
Risky Genetic	0.012	0.019	-0.001	0.018
Score	[0.010]	[0.009]**	[0.007]	[0.006]***
Mom Risky Gene	0.000	-0.012	0.005	0.008
	[0.010]	[0.009]	[0.007]	[0.006]
$log(BMI_{t-1})$	-0.105	-0.161	0.078	0.045
	[0.036]***	[0.032]***	[0.025]***	[0.021]**
Controls	X	X	X	X
Observations	3,346	3,706	3,346	3,706

* significant at 10%; ** significant at 5%; *** significant at 1%. Standard error clustered at the individual level in brackets. Dependent variable: logarithm of daily kilocalories intake and daily sedentary minutes; Controls: 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; late respondent; birth weight.

To summarize, I find that the FTO gene 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.

Although these results are consistent with the evidence from the genetic literature

on FTO, a replication using a different dataset is warranted, as suggested by Benjamin et al. (2011).

4 Conclusion

I introduce a general economic framework that combines the recent discoveries of molecular genetics with a model of human capital formation in the early periods of life. This enables us to understand how family decisions about investments in the human capital of the child are affected by his genetic endowment. I show how genes interact with the environment in two ways: they can shift the implicit costs of investments; and they can change the productivity of inputs in the formation of human capital. These genetic effects induce a change in the optimal allocation of family resources, and in the equilibrium level of human capital. The model is general enough to be applied to different types of human capital, investments, and genes.

I test the model focusing on obesity, a negative feature of health and human capital that increased alarmingly in the last decades. 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 leverage recent findings in human genetics that show how the FTO gene influences appetite and food satiation.

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. I find evidence of a productivity effect of the FTO gene, 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, children endowed with at least one A-allele in the rs9939609 FTO gene variant still end up ingesting more calories. This is consistent with a genetic effect on the cost of investment in diet. There is no prominent genetic effect connected to physical activity. To summarize, the FTO gene is associated with children who eat more as well as a higher risk of obesity for those who ingest a lot of calories. These two effects jointly explain why children with this particular genotype tend to be more obese. However this result is conditional on the 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, and different estimation strategies. Using different measurements of obesity or investments does not change the qualitative findings. I also find interesting gender differences: on average, males tend to be more active and have a higher caloric intake. More interestingly, the genetic productivity effects are gender specific: the FTO gene increases the productivity of investment in exercise for males, and investment in diet for females. In other words, I find evidence of a gene-diet interaction for females, and gene-exercise interaction for males. 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.

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 40-60% of the variation in obesity has been estimated to be heritable and due to genetic endowment, 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.

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.

References

Anastasi, A. (1958). Heredity, environment, and the question "How?". *Psychological Review*, 65:197–208.

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

Becker, G. S. (2007b). Health as human capital: synthesis and extensions. *Oxford Economic Papers*, 59(3):379–410.

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., Cesarini, D., Chabris, C. F., Glaeser, E. L., Laibson, D. I., Gunason, V., Harris, T. B., Launer, L. J., Purcell, S., Smith, A. V., Johannesson, M., Magnusson, P. K., Beauchamp, J. P., Christakis, N. A., Atwood, C. S., Hebert, B., Freese, J., Hauser, R. M., Hauser, T. S., Grankvist, A., Hultman, C. M., and Lichtenstein, P. (2011). The Promises and Pitfalls of Genoeconomics. *Annual Review of Economics*, 4(1).

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.

Butcher, L. M., Kennedy, J. K., and Plomin, R. (2006). Generalist genes and cognitive neuroscience. *Current opinion in neurobiology*, 16(2):145–51.

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. 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.

Cole, T. J., Bellizzi, M. C., Flegal, K. M., and Dietz, W. H. (2000). Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ*, 320(1240):1–6.

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.

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.

Cosconati, M. (2009). Parenting style and the development of human capital in children. *Job Market Paper, University of Pennsylvania*.

Cosconati, M. (2012). Optimal Parenting Styles: Evidence From a Dynamic Model with Multiple Equilibria. *Working Paper*.

Cunha, F. and Heckman, J. J. (2007). The Technology of Skill Formation. *American Economic Review*, 97(2):31–47.

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.

Dauphin, A., El Lahga, A.-R., Fortin, B., and Lacroix, G. (2011). Are Children Decision-Makers within the Household? *The Economic Journal*, 121:871–903.

Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D., Ke, X., Le Hellard, S., Christoforou, A., Luciano, M., McGhee, K., Lopez, L., Gow, A. J., Corley, J., Redmond, P., Fox, H. C., Haggarty, P., Whalley, L. J., McNeill, G., Goddard, M. E., Espeseth, T., Lundervold, A. J., Reinvang, I., Pickles, A., Steen, V. M., Ollier, W., Porteous, D. J., Horan, M., Starr, J. M., Pendleton, N., Visscher, P. M., and Deary, I. J. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular psychiatry*, 16(10):996–1005.

Dickens, W. T. and Flynn, J. R. (2001). Heritability estimates versus large environmental effects: the IQ paradox resolved. *Psychological review*, 108(2):346–69.

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.

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.

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., 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.

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., Hullegie, P., Erik, M., and Outcault, S. (2012). Empirical evidence for decreasing returns to scale in a health capital model. *Working Paper*.

Galton, F. (1874). *English men of science: Their nature and nurture*. Number v. 27; v. 469 in Popular science library. D. Appleton.

Goldberger, A. (1976). *Twin methods: A skeptical view*. Social Systems Research Institute, University of Wisconsin-Madison.

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

Golding, J., Pembrey, M., Jones, R., and ALSPAC Study Team, T. (2001). ALSPAC–the Avon Longitudinal Study of Parents and Children. I. Study methodology. *Paediatric and perinatal epidemiology*, 15(1):74–87.

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.

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.

Heckman, J. J. (2012). The developmental origins of health. *Health Economics*, 29:24–29.

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, L., Burn, P., and Lee, F. (1997). Targeted Disruption of the Melanocortin-4 Receptor Results in Obesity in Mice. *Cell*, 88(1):131–141.

Kilpeläinen, T. O., Qi, L., Brage, S., Sharp, S. J., Sonestedt, E., Demerath, E., 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., Grø ntved, 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., Sø rensen, T. I. A., Mohlke, K. L., Hofman, A., Uitterlinden, A. G., Tuomilehto, J., Lehtimäki, T., Raitakari, O., Isomaa, B., Njø lstad, P. l. 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.

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.

Lundberg, S., Romich, J. L., and Tsang, K. P. (2009). Decision-making by children. *Review of Economics of the Household*, 7(1):1–30.

Manski, C. F. (2011). Genes, Eyeglasses, and Social Policy. *Journal of Economic Perspectives*.

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.

Meaney, M. J. (2010). Epigenetics and the biological definition of gene x environment interactions. *Child development*, 81(1):41–79.

Meaney, M. J. and Szyf, M. (2005). Environmental programming of stress responses through DNA methylation: life at the interface between a dynamic environment and a fixed genome. *Dialogues in Clinical Neuroscience*, 3(7):103–123.

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.

Moffitt, T. E., Caspi, A., and Rutter, M. (2006). Measured Gene-Environment Interactions in Psychopathology Concepts, Research Strategies, and Implications for Research, Intervention, and Public Understanding. *Perspectives on Psychological Science*, 1(1):5–27.

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

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 : the journal of the American Medical Association*, 307(5):483–90.

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.

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. C., 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*, 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.

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.

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.

Silventoinen, K., Hasselbalch, A. L., Lallukka, T., Bogl, L., Pietila, K. H., Heitmann, B. L., Schousboe, K., Rissanen, A., Kyvik, K. O., Sø rensen, 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 : an official journal of the International Association for the Study of Obesity*, 9(5):474–88.

Smith, A. (1776). *An Inquiry into the Nature and Causes of The Wealth of Nations*. W. Strahan and T. Cadell, London.

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., Loos, R. J. F., , et al. (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.

Szyf, M. and Bick, J. (2013). DNA Methylation: A Mechanism for Embedding Early Life Experiences in the Genome. *Child development*, 84(1):49–57.

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.

van der Sluis, S., Willemsen, G., de Geus, E. J. C., Boomsma, D. I., and Posthuma, D. (2008). Gene-environment interaction in adults' IQ scores: measures of past and present environment. *Behavior genetics*, 38(4):348–60.

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.

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.

West, M. J. and King, A. P. (1987). Settling nature and nurture into an ontogenetic niche. *Developmental psychobiology*, 20(5):549–62.

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, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A. F., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E., and Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature genetics*, 42(7):565–9.

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

Yeo, G. S. H. and Heisler, L. K. (2012). Unraveling the brain regulation of appetite: lessons from genetics. *Nature neuroscience*, 15(10):1343–9.

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

Appendices

A The Model

In section $(A.1)$ I develop the simplest version of the model, where the the family chooses directly the investment levels to maximize the human capital of the child, similarly to a firm's production problem. This simpler problem can be viewed as nested into a home production model, where the family decides how much time and goods to invest in the human capital of the child, and how much to consume. This more general model is developed in section$(A.2)$, where I introduce investment functions that depend on the allocation of goods and time, in the spirit of Grossman (1972). I then derive how the (shadow) prices of the investments depend on the genetic endowment of the child and the prices of goods and time.

A.1 Human Capital Production

First let's look at the simpler model where food consumption and exercise are considered directly as inputs, whose price varies with genetic endowment.

$$\begin{aligned} \max_{I_d, I_e} & H_t = f(I_d, I_e; g) + (1 - \delta(g)) H_{t-1} \\ s.t & Y = p_d(g) I_d + p_e(g) I_e \end{aligned}$$

H is the stock of health (or human capital), which is function of different types of investments - in this case I consider two potential investment, exercise I_e and diet I_d. The parameters of the production function $H(.)$ are indexed by the child's genotype g. Y are the family resources dedicated to the investment in the child's human capital, and split between I_e and I_d according to their relative prices $p_x(g)$ for $k = e, d$, which depend on the genetic endowment of the individual.

This is an important feature of the model: genes enter the production function of human capital, but also the price of investment. A different genetic makeup will induce variations in the subjective cost of investments.[48]

The associated Lagrangian is $\mathcal{L} = H(I_d, I_e; g) + \lambda_y (Y - p_d(g) I_d - p_e(g) I_e)$. To find the solution of this model, maximize with respect to the choice variables I_d, I_e. The following first order conditions are obtained:

$$\frac{\partial H(I_d, I_e; g)}{\partial I_e} = \lambda_y p_e(g)$$

[48] See the discussion in the text for the molecular genetics bases of this assumption as well as a more general model that derives the shadow prices of these two inputs from a more basic investment function that depends on genes, goods, and time and effort.

$$\frac{\partial H\left(I_d, I_e; g\right)}{\partial I_d} = \lambda_y p_d(g)$$

Which leads to the usual condition of marginal rates of substitution equal to the ratio of prices:

$$\frac{\frac{\partial H(I_d, I_e; g)}{\partial I_e}}{\frac{\partial H(I_d, I_e; g)}{\partial I_d}} = \frac{p_e(g)}{p_d(g)}$$

A.1.1 Functional form Specification

To find a closed form solution to the model, consider different specifications for the the production function of health: for example a Cobb-Douglas specification or a more general Constant Elasticity of Substitution (CES).

Cobb-Douglas First, let's assume a Cobb-Douglas functional form specification also for the production function of health: $f\left(I_d, I_e; g\right) = A(g)\left(I_e^{\alpha_e(g)} I_d^{\alpha_d(g)}\right)$. For generality, allow the parameters of the production function to be dependent on the genotype of the child. The problem becomes:

$$\max_{I_d, I_e} \quad \left[\log A(g) + \alpha_e(g) \log I_e + \alpha_d(g) \log I_d\right] + \lambda_y \left(Y - p_d(g) I_d - p_e(g) I_e\right)$$

The first order condition become:

$$\begin{aligned} \frac{\alpha_d(g)}{I_d^*} &= \lambda_y p_d(g) \\ \frac{\alpha_e(g)}{I_e^*} &= \lambda_y p_e(g) \end{aligned}$$

Solving as a function of λ_y leads to:

$$\lambda_y = \frac{\alpha_e(g)}{I_e^* p_e(g)} = \frac{\alpha_d(g)}{I_d^* p_d(g)}$$

So that the relation between the investments is:

$$I_e^* = \frac{p_d(g)}{p_e(g)} \frac{\alpha_e(g)}{\alpha_d(g)} I_d^*$$

Substitute into the budget constraint, in order to obtain the demand as function of prices and income:

$$Y = p_e(g) I_e^* + p_d(g) I_d^*$$

$$
\begin{aligned}
&= p_e(g)\frac{\alpha_e(g)}{p_e(g)}\frac{p_d(g)}{\alpha_d(g)}I_d^* + p_d(g)I_d^* \\
&= \left(\frac{\alpha_e(g)}{\alpha_d(g)} + 1\right) p_d(g)I_d^*
\end{aligned}
$$

in order to obtain:

$$
\begin{aligned}
I_d^* &= \frac{\alpha_d(g)}{p_d(g)}\frac{1}{\alpha_e(g)+\alpha_d(g)}Y \\
I_e^* &= \frac{\alpha_e(g)}{p_e(g)}\frac{1}{\alpha_e(g)+\alpha_d(g)}Y
\end{aligned}
$$

Note that the optimal level of investment is inversely related to is price $p_k(g)$ and proportional to the productivity of investment α_k. If genes increase the productivity of one input ($\alpha_k(g=1) > \alpha_k(g=0)$) , we expect this input to increase, while if they increase its price ($\alpha_k(g=1) > \alpha_k(g=0)$)) we expect the equilibrium level of I_k^* to decrease. Finally, notice that if genes have an overall productivity effect, so that $A(g=1) > A(g=0)$, this would lead to an overall higher level of human capital, but there would be no changes in the optimal level of either investments or consumption. [49]

CES A similar result holds when considering the case of Constant Elasticity of Substitution, so that the production function of health becomes:

$f\left(I_d, I_e; g\right) = A(g)\left[\alpha_d(g)I_d^{\eta(g)} + (1-\alpha_d(g))I_e^{\eta(g)}\right]^{\frac{\nu}{\eta(g)}}$

As $\eta(g) \to 0$ we have that the CES simplifies into a Cobb-Douglas, with weights $\nu\alpha_d(g)$ and $\nu(1-\alpha_d(g))$.

The family problem becomes:

$$
\begin{aligned}
\max_{I_d, I_e, c} \quad & \left[\log A(g) + \frac{\nu}{\eta(g)}\log\left[\alpha_d(g)I_d^{\eta(g)} + (1-\alpha_d(g))I_e^{\eta(g)}\right]\right] \\
s.t. \quad & Y = p_d(g)I_d + p_e(g)I_e
\end{aligned}
$$

For notational simplicity, let's remove the explicit genetic dependence (g), remembering that both prices and the parameters can depend on the child's genotype. The first order condition become:

$$
\frac{\partial \mathcal{L}}{\partial I_e} = \frac{(1-\alpha_d)I_e^{*\eta-1}}{\left[\alpha_d I_d^{*\eta} + (1-\alpha_d)I_e^{*\eta}\right]}\frac{\nu\eta}{\eta} - \lambda_y p_e = 0
$$

[49] Note that if it is assumed that $\alpha_e(g) + \alpha_d(g) = 1$, so that there are constant returns to scale to I_e and I_d in the production function of health, then the results simplifies to the usual result that the optimal consumption of c and H will be proportional to income, with weights determined by their importance in the utility function (ϕ and $1-\phi$ respectively) and inversely proportional to their prices; for example $c^* = \phi Y$. However if a more general specification is allowed for the utility function with substitution between consumption and human capital, so that $\frac{\partial U}{\partial x \partial H} \neq 0$, then an increase in genetic productivity would have effects on the consumption as well.

$$\frac{\partial \mathcal{L}}{\partial I_d} = \frac{\alpha_d I_d^{*\eta-1}}{[\alpha_d I_d^{*\eta} + (1-\alpha_d) I_e^{*\eta}]} \frac{\nu\eta}{\eta} - \lambda_y p_d = 0$$

Or equivalently, expressed in terms of λ_y:

$$\begin{aligned} \lambda_y &= \frac{1}{p_e} \frac{\nu(1-\alpha_d) I_e^{*\eta-1}}{f\left(I_e^*, I_d^*; g\right)^{\eta/\nu} / A} \\ &= \frac{1}{p_d} \frac{\nu\alpha_d I_d^{*\eta-1}}{f\left(I_e^*, I_d^*; g\right)^{\eta/\nu} / A} \end{aligned}$$

The choices of substitution between the two inputs becomes:

$$\begin{aligned} \frac{p_d}{p_e} &= \frac{\alpha_d}{1-\alpha_d} \left(\frac{I_d^*}{I_e^*}\right)^{\eta-1} \\ &= \frac{\alpha_d}{1-\alpha_d} \left(\frac{I_e^*}{I_d^*}\right)^{1-\eta} \\ I_e^* &= \left[\frac{1-\alpha_d}{\alpha_d} \frac{p_d}{p_e}\right]^{1/1-\eta} I_d^* \end{aligned}$$

Even in this case we have that the ratio of derivatives depend both on prices and the relative efficiency of the two inputs; furthermore, also the elasticity of substitution η plays a role: it increases the sensitivity to changes in prices and productivity if $\eta < 0$ (the elasticity $\sigma = \frac{1}{1-\eta} > 1$), and decreases it otherwise.

Substituting everything into the budget constraint we obtain:

$$\begin{aligned} Y &= p_d I_d^* + p_e I_e^* \\ &= p_d I_d^* + p_e \left[\frac{1-\alpha_d}{\alpha_d} \frac{p_d}{p_e}\right]^{1/1-\eta} I_d^* \\ &= I_d^* \left[p_d + \left(\frac{p_d}{p_e^\eta} \frac{1-\alpha_d}{\alpha_d}\right)^{1/1-\eta}\right] \\ &= I_d^* p_d \left[1 + \left(\frac{p_d}{p_e}\right)^{\eta/1-\eta} \left(\frac{1-\alpha_d}{\alpha_d}\right)^{1/1-\eta}\right] \end{aligned}$$

Let's use $B = \left(\frac{1-\alpha_d}{\alpha_d}\right)^{1/1-\eta} \left(\frac{p_d}{p_e}\right)^{\eta/1-\eta}$. Note that as $\eta \to 0$ $B \to \frac{1-\alpha_d}{\alpha_d}$, $\frac{1}{1+B} \to \alpha_d$ and $\frac{B}{1+B} \to 1-\alpha_d$. We have that

$$\begin{aligned} I_d^* &= \frac{1}{p_d} \frac{1}{1+B} Y \\ I_e^* &= \frac{1}{p_e} \frac{B}{1+B} Y \end{aligned}$$

A.2 Household Production Model With Goods and Time Inputs

In order to give some micro foundations to the claim that the prices of investment can depend on the genetic makeup of an individual, let's consider the following household production model:

$$\begin{aligned} \max_{\{x_k,\tau_k\}_{k=e}^{d},c} \quad & U(c,\tau_l,H) \\ s.t. & \\ \Omega &= \tau_l+\tau_d+\tau_e \\ Y &= p_c c+p_d x_d+p_e x_e \\ H &= f\left(I_d,I_e;g\right) \\ I_d &= I_d\left(x_d,\tau_d;g\right) \\ I_e &= I_e\left(x_e,\tau_e;g\right) \end{aligned} \tag{6}$$

Again, the parameters of the production function $f(.)$ of human capital are indexed by the genotype of the child g. The investment levels I_k, for $k \in (e,d)$, are obtained combining goods x_k and effort τ_k; the relative productivity of these two inputs might depend on the genetic endowment of the child, and therefore the functions $I_k\left(x_k,\tau_k;g\right)$ are indexed by g. Ω is the total amount of time and energy that the family can choose to allocate to leisure τ_l, or investing in the child's human capital τ_k. Y is income, which can be allocated to either consumption c or investments x, according to their relative prices p_k. The two investments I_k are themselves function of market goods, whose prices are equal for everybody, as well as time and effort. The genetic endowment of the individual interact with these inputs, changing their relative productivity.

To find a solution to the model, consider the Lagrangian associated to this maximization and substitute all of the investment functions into the main production function of human capital

$$\begin{aligned} \mathcal{L} &= U\left[c,\tau_l,H\right]+\lambda_\tau\left(\Omega-\tau_l-\tau_e-\tau_d\right)+\lambda_y\left(Y-p_c c-p_d x_d-p_e x_e\right) \\ &= U\left[c,\tau_l,f\left(I_e\left(x_e,\tau_e;g\right),I_d\left(x_d,\tau_d;g\right);g\right)\right]+ \\ &\quad +\lambda_\tau\left(\Omega-\tau_l-\tau_e-\tau_d\right)+\lambda_y\left(Y-p_c c-p_d x_d-p_e x_e\right) \end{aligned}$$

Assuming that all of the time not used investing is devoted to leisure, the time constraint is always binding and $\tau_l=\Omega-\tau_e-\tau_d$ so that

$$\begin{aligned} \mathcal{L} &= U\left[c,\Omega-\tau_e-\tau_d,f\left(I_e\left(x_e,\tau_e;g\right),I_d\left(x_d,\tau_d;g\right);g\right)\right]+ \\ &\quad +\lambda_y\left(Y-p_c c-p_d x_d-p_e x_e\right) \end{aligned}$$

Maximizing with respect to the goods c,x_e,x_d and time τ_e,τ_d, and normalizing for the price of the consumption good ($p_c=1$), leads to the following first order conditions:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial c} &= \frac{\partial U}{\partial c} - \lambda_y = 0 \\
\frac{\partial \mathcal{L}}{\partial x_e} &= \frac{\partial U}{\partial H}\frac{\partial f\left(I_d, I_e; g\right)}{\partial I_e\left(x_e, \tau_e; g\right)}\frac{\partial I_e\left(x_e, \tau_e; g\right)}{\partial x_e} - \lambda_y p_e = 0 \\
\frac{\partial \mathcal{L}}{\partial x_d} &= \frac{\partial U}{\partial H}\frac{\partial f\left(I_d, I_e; g\right)}{\partial I_d\left(x_d, \tau_d; g\right)}\frac{\partial I_d\left(x_d, \tau_d; g\right)}{\partial x_d} - \lambda_y p_d = 0
\end{aligned}$$

And

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \tau_e} &= \frac{\partial U}{\partial H}\frac{\partial f\left(I_d, I_e; g\right)}{\partial I_e\left(x_e, \tau_e; g\right)}\frac{\partial I_e\left(x_e, \tau_e; g\right)}{\partial \tau_e} - \frac{\partial U}{\partial \tau_l} = 0 \\
\frac{\partial \mathcal{L}}{\partial \tau_d} &= \frac{\partial U}{\partial H}\frac{\partial f\left(I_d, I_e; g\right)}{\partial I_d\left(x_d, \tau_d; g\right)}\frac{\partial I_d\left(x_d, \tau_d; g\right)}{\partial \tau_d} - \frac{\partial U}{\partial \tau_l} = 0
\end{aligned}$$

The first set of equations show that the optimal level of goods spent in investment (x_e^*, x_d^*) is such that the marginal productivity of every dollar spent on investment has to be equal to the marginal utility of every dollar spent on consumption.[50]

$$\frac{\partial U}{\partial H}\frac{\partial f\left(I_d, I_e; g\right)}{\partial I_e\left(x_e, \tau_e; g\right)}\frac{\partial I_e\left(x_e, \tau_e; g\right)}{\partial x_e}\frac{1}{p_e} = \frac{\partial U}{\partial H}\frac{\partial f\left(I_d, I_e; g\right)}{\partial I_d\left(x_d, \tau_d; g\right)}\frac{\partial I_d\left(x_d, \tau_d; g\right)}{\partial x_d}\frac{1}{p_d} = \frac{\partial U}{\partial c}$$

Focusing on the investment part leads to:

$$\frac{\frac{\partial f(I_d, I_e; g)}{\partial I_e(x_e, \tau_e; g)}\frac{\partial I_e(x_e, \tau_e; g)}{\partial x_e}}{p_e} = \frac{\frac{\partial f(I_d, I_e; g)}{\partial I_d(x_d, \tau_d; g)}\frac{\partial I_d(x_d, \tau_d; g)}{\partial x_d}}{p_d}$$

Or equivalently:

$$\frac{\frac{\partial f(I_d, I_e; g)}{\partial I_e(x_e, \tau_e; g)}\frac{\partial I_e(x_e, \tau_e; g)}{\partial x_e}}{\frac{\partial f(I_d, I_e; g)}{\partial I_d(x_d, \tau_d; g)}\frac{\partial I_d(x_d, \tau_d; g)}{\partial x_d}} = \frac{p_e}{p_d}$$

Rearrange the terms to obtain:

$$\begin{aligned}
\frac{\frac{\partial f(I_d, I_e; g)}{\partial I_e(x_e, \tau_e; g)}}{\frac{\partial f(I_d, I_e; g)}{\partial I_d(x_d, \tau_d; g)}} &= \frac{p_e / \frac{\partial I_e(x_e, \tau_e; g)}{\partial x_e}}{p_d / \frac{\partial I_d(x_d, \tau_d; g)}{\partial x_d}} \\
\frac{f'_{I_e}}{f'_d} &= \frac{p_e / I'_e}{p_d / I'_d} = \frac{p'_e}{p'_d}
\end{aligned}$$

[50] If the genetic endowment changed preferences and entered the utility function, then these margins would be different depending on the agent's genotype.

A.2.1 Functional form Specification

Let's make some functional form specifications that will ease the solution of the model. For simplicity, let's assume an additive separable utility function and a Cobb-Douglas specification for all of the production functions.

$$\begin{aligned}
\max_{x_e,x_d,\{\tau_k\}_{k=e}^{d},c} \quad & \phi_1 \log c + \phi_2 \log \tau_l + \phi_3 \log H \\
s.t. & \\
\Omega &= \tau_l + \tau_e + \tau_d \\
Y &= p_c c + p_d x_d + p_e x_e \\
H &= A_H(g)\left(I_e^{\alpha_e} I_d^{\alpha_d}\right) \\
I_e &= A_e(g)\left(x_e^{\gamma_1}\tau_e^{\gamma_2}\right) \\
I_d &= A_d(g)\left(x_d^{\delta_1}\tau_d^{\delta_2}\right)
\end{aligned}$$

As before, consider that all the parameters of the functions can be dependent on genes. However this time the prices of the goods x_k and the time τ_k do not depend on genes. Assuming that all of the time not used investing is devoted to leisure, the time constraint is always binding and $\tau_l = \Omega - \tau_e - \tau_d$ so that

$$\begin{aligned}
\mathcal{L} &= \phi_1 \log c + \phi_2 \log\left(\Omega - \tau_e - \tau_d\right) \\
&\quad + \phi_3 \log A_H(g)\left[\left(A_e(g)x_e^{\gamma_1}\tau_e^{\gamma_2}\right)^{\alpha_e}\left(A_d(g)x_d^{\delta_1}\tau_d^{\delta_2}\right)^{\alpha_d}\right] + \\
&\quad + \lambda_y\left(Y - p_c c - p_d x_d - p_e x_e\right) \\
&= \phi_1 \log c + \phi_2 \log\left(\Omega - \tau_e - \tau_d\right) \\
&\quad + \phi_3 \log A_H(g) A_e(g) A_d^{\alpha_d}(g) \\
&\quad + \phi_3\gamma_1\alpha_e \log x_e + \phi_3\delta_1\alpha_d \log x_d + \\
&\quad + \phi_3\gamma_2\alpha_e \log \tau_e + \phi_3\delta_2\alpha_d \log \tau_d + \\
&\quad + \lambda_y\left(Y - p_c c - p_d x_d - p_e x_e\right)
\end{aligned}$$

It is easy to see that g enter in two main ways. Firstly as a constant, similar to the effect of A, increasing the overall productivity of investments and shifting outward the productivity frontier: for the same combination of investment, a higher human capital is obtained, effectively reducing the overall cost of the child's achievement. Secondly by changing the parameters of the functions.

Maximizing with respect to the goods c, x_e, x_d and time τ_e, τ_d, and normalizing for the price of the consumption good ($p_c = 1$), leads to the following first order conditions:

$$\frac{\partial \mathcal{L}}{\partial c} = \frac{\phi_1}{c} - \lambda_y = 0$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial x_e} &= \frac{\phi_3 \gamma_1 \alpha_e}{x_e} - \lambda_y p_e = 0 \\ \frac{\partial \mathcal{L}}{\partial x_d} &= \frac{\phi_3 \delta_1 \alpha_d}{x_d} - \lambda_y p_d = 0\end{aligned}$$

And

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \tau_e} &= \frac{\phi_3 \gamma_2 \alpha_e}{\tau_e} - \frac{\phi_2}{\Omega - \tau_e - \tau_d} = 0 \\ \frac{\partial \mathcal{L}}{\partial \tau_d} &= \frac{\phi_3 \delta_2 \alpha_d}{\tau_d} - \frac{\phi_2}{\Omega - \tau_e - \tau_d} = 0\end{aligned}$$

Focusing on the optimal allocation of goods:

$$\begin{aligned}\frac{\phi_1}{c} &= \frac{\phi_3 \gamma_1 \alpha_e}{p_e x_e} = \frac{\phi_3 \delta_1 \alpha_d}{x_d p_d} \\ x_e &= \frac{\phi_3}{\phi_1} \frac{\gamma_1 \alpha_e}{p_e} c \\ x_d &= \frac{\phi_3}{\phi_1} \frac{\delta_1 \alpha_d}{p_d} c\end{aligned}$$

Substitute into the budget constraint to obtain:

$$\begin{aligned}Y &= c + p_e x_e + p_d x_d \\ &= c \left[1 + p_e \frac{\phi_3}{\phi_1} \frac{\gamma_1 \alpha_e}{p_e} + p_d \frac{\phi_3}{\phi_1} \frac{\delta_1 \alpha_d}{p_d} \right] \\ &= c \left[\frac{\phi_1 + \phi_3 \left(\gamma_1 \alpha_e + \delta_1 \alpha_d \right)}{\phi_1} \right]\end{aligned}$$

so that the optimal allocations become:

$$\begin{aligned}c^* &= \frac{\phi_1}{\phi_1 + \phi_3 \left(\gamma_1 \alpha_e + \delta_1 \alpha_d \right)} Y \\ x_e^* &= \frac{\phi_3 \gamma_1 \alpha_e}{\phi_1 + \phi_3 \left(\gamma_1 \alpha_e + \delta_1 \alpha_d \right)} \frac{Y}{p_e} \\ x_d^* &= \frac{\phi_3 \delta_1 \alpha_d}{\phi_1 + \phi_3 \left(\gamma_1 \alpha_e + \delta_1 \alpha_d \right)} \frac{Y}{p_d}\end{aligned}$$

Therefore the optimal amount of inputs will be a function of the genes if the parameters α, δ, γ vary with the genotype of the child.

Looking at the FOC for time allocation, the ratio of the time spent investing in the two inputs is equal to the relation of productivity coefficients: $\frac{\tau_d}{\tau_e} = \frac{\delta_2 \alpha_d}{\gamma_2 \alpha_e}$. Using all of the time available leads to $\Omega - \tau_e - \tau_d = \tau_l$, so that $\frac{\tau_l}{\tau_e} = \frac{\phi_3 \gamma_2 \alpha_e}{\phi_2}$. Substituting into the time budget constrain:

$$\Omega = \tau_l^* + \tau_e^* + \tau_d^*$$

$$= \left[\frac{\phi_3\gamma_2\alpha_e}{\phi_2} + 1 + \frac{\delta_2\alpha_d}{\gamma_2\alpha_e}\right]\tau_e^*$$

so that the optimal allocation becomes:

$$\tau_e^* = \frac{\phi_2\gamma_2\alpha_e}{\phi_3\gamma_2^2\alpha_e + \phi_2\gamma_2\alpha_e + \phi_2\delta_2\alpha_d}\Omega$$
$$\tau_d^* = \frac{\phi_2\delta_2\alpha_d}{\phi_3\gamma_2^2\alpha_e + \phi_2\gamma_2\alpha_e + \phi_2\delta_2\alpha_d}\Omega$$
$$\tau_l^* = \frac{\phi_3\gamma_2^2\alpha_e}{\phi_3\gamma_2^2\alpha_e + \phi_2\gamma_2\alpha_e + \phi_2\delta_2\alpha_d}\Omega$$

Therefore the optimal investment will be $I_k^* = A_k(g)\left(x_k^{*\gamma_1}\tau_k^{*\gamma_2}\right)$, a function of full income (Y, Ω), the prices p_k of the goods, and the genotype-specific parameters of the model. The shadow price of investment will be $p_k' = p_k x_k^* + w\tau_k^*$, where w is the shadow price of time. p_k' will depend on the genetic endowment of the child, since the optimal level of goods $x_k^*$and time devoted to investment τ_k^* will be.

CES for the Investment Functions Let's assume instead that the investment functions follow a Constant Elasticity of Substitution functional form, while the production function a normal Cobb-Douglas. The family maximizes the following:

$$\max_{x_e, x_d, \{\tau_k\}_{k=e}^d, c} \phi_1 \log c + \phi_2 \log \tau_l + \phi_3 \log H$$
$$s.t.$$
$$\Omega = \tau_l + \tau_e + \tau_d$$
$$Y = p_c c + p_e x_e + p_d x_d$$
$$H = A_H\left(I_e^{\alpha_e} I_d^{\alpha_d} g^{\alpha_g}\right)$$
$$I_e\left(x_e, \tau_e; g\right) = A_e\left[\gamma_1 x_e^{\eta_e} + \gamma_2 \tau_e^{\eta_e} + \left(1 - \gamma_1 - \gamma_2\right) g^{\eta_e}\right]^{\frac{1}{\eta_e}}$$
$$I_d\left(x_d, \tau_d; g\right) = A_d\left[\delta_1 x_d^{\eta_d} + \delta_2 \tau_d^{\eta_d} + \left(1 - \delta_1 - \delta_2\right) g^{\eta_d}\right]^{\frac{1}{\eta_d}}$$

This specification leads to the following derivatives:

$$\frac{\partial I_e\left(x_e, \tau_e; g\right)}{\partial x_e} = \gamma_1 x_e^{\eta_e - 1}\frac{\eta_e}{\eta_e} A_e\left[\gamma_1 x_e^{\eta_e} + \gamma_2 \tau_e^{\eta_e} + \left(1 - \gamma_1 - \gamma_2\right) g^{\eta_e}\right]^{\frac{1}{\eta_e} - 1}$$
$$= \gamma_1 x_e^{\eta_e - 1} I_e^{1 - \eta_e} A_e^{\eta_e}$$
$$\frac{\partial I_d\left(x_d, \tau_d; g\right)}{\partial x_d} = \delta_1 x_d^{\eta_d - 1}\frac{\eta_d}{\eta_d} A_d\left[\delta_1 x_d^{\eta_d} + \delta_2 \tau_d^{\eta_d} + \left(1 - \delta_1 - \delta_2\right) g^{\eta_d}\right]^{\frac{1}{\eta_d} - 1}$$
$$= \delta_1 x_d^{\eta_d - 1} I_d^{1 - \eta_d} A_d^{\eta_d}$$

Focusing only on the investment decisions, the following must hold:

$$\frac{\frac{\partial f(I_d, I_e; g)}{\partial I_e(x_e, \tau_e; g)}}{\frac{\partial f(I_d, I_e; g)}{\partial I_d(x_d, \tau_d; g)}} = \frac{p_e / \frac{\partial I_e(x_e, \tau_e; g)}{\partial x_e}}{p_d / \frac{\partial I_d(x_d, \tau_d; g)}{\partial x_d}}$$

$$
\begin{aligned}
\frac{\alpha_e H I_e^{-1}}{\alpha_d H I_d^{-1}} &= \frac{p_e/\gamma_1 x_e^{\eta_e-1} I_e^{1-\eta_e} A_e^{\eta_e}}{p_d/\delta_1 x_d^{\eta_d-1} I_d^{1-\eta_d} A_d^{\eta_d}} \\
\frac{\alpha_e I_e^{-1} x_e^{\eta_e-1} I_e^{1-\eta_e}}{\alpha_d I_d^{-1} x_d^{\eta_d-1} I_d^{1-\eta_d}} &= \frac{p_e/\gamma_1 A_e^{\eta_e}}{p_d/\delta_1 A_d^{\eta_d}} \\
\frac{x_e^{\eta_e-1}/\left(A_e I_e\right)^{\eta_e}}{x_d^{\eta_d-1}/\left(A_d I_d\right)^{\eta_d}} &= \frac{p_e/\gamma_1 \alpha_e}{p_d/\delta_1 \alpha_d} \\
\frac{x_e^{\eta_e-1}/\left[\gamma_1 x_e^{\eta_e}+\gamma_2 \tau_e^{\eta_e}+\left(1-\gamma_1-\gamma_2\right) g^{\eta_e}\right]}{x_d^{\eta_d-1}/\left[\delta_1 x_d^{\eta_d}+\delta_2 \tau_d^{\eta_d}+\left(1-\delta_1-\delta_2\right) g^{\eta_d}\right]} &= \frac{p_e/\gamma_1 \alpha_e}{p_d/\delta_1 \alpha_d}
\end{aligned}
$$

And, looking at the time allocation, a similar result is found:

$$
\begin{aligned}
\frac{\partial f\left(I_d, I_e; g\right)}{\partial I_e\left(x_e, \tau_e; g\right)} \frac{\partial I_e\left(x_e, \tau_e; g\right)}{\partial \tau_e} &= \frac{\partial f\left(I_d, I_e; g\right)}{\partial I_d\left(x_d, \tau_d; g\right)} \frac{\partial I_d\left(x_d, \tau_d; g\right)}{\partial \tau_d} \\
\alpha_e H I_e^{-1} \gamma_2 A_e^{\eta_e} I_e^{1-\eta_e} \tau_e^{\eta_e-1} &= \alpha_d H I_d^{-1} \delta_2 A_d^{\eta_d} I_d^{1-\eta_d} \tau_d^{\eta_d-1} \\
\frac{\tau_e^{\eta_e-1}/\left(A_e I_e\right)^{\eta_e}}{\tau_d^{\eta_d-1}/\left(A_d I_d\right)^{\eta_d}} &= \frac{\alpha_d \delta_2}{\alpha_e \gamma_2}
\end{aligned}
$$

Therefore the optimal level of goods $(x_e, x_d)^*$ and time $(\tau_e, \tau_d)^*$ will depend on the genetic makeup, and their ratio will be a function of the prices and the productivity:

$$
\begin{aligned}
\frac{\tau_e^{1-\eta_e}}{\tau_d^{1-\eta_d}} \frac{\alpha_d \delta_2}{\alpha_e \gamma_2} &= \frac{p_e/\gamma_1 \alpha_e}{p_d/\delta_1 \alpha_d} \frac{x_e^{1-\eta_e}}{x_d^{1-\eta_d}} \\
\frac{\left(x_e/\tau_e\right)^{1-\eta_e}}{\left(x_d/\tau_d\right)^{1-\eta_d}} &= \frac{p_d}{p_e} \frac{\gamma_1 \delta_2}{\gamma_2 \delta_1} \\
\frac{x_e}{\tau_e} &= \left(\frac{p_d}{p_e} \frac{\gamma_1 \delta_2}{\gamma_2 \delta_1}\right)^{1/1-\eta_e} \left(\frac{x_d}{\tau_d}\right)^{\frac{1-\eta_d}{1-\eta_e}}
\end{aligned}
$$

B Likelihood

In section (3.4) we estimate a Constant Elasticity of Substitution production function. We have the following specification for $\log(H_{i,t})$, the logarithm of Body-Mass-Index of individual i at time t:

$$
\log H_{i,t} = \frac{\nu}{\phi} \log\left[\alpha {I_{i,t}^e}^{\phi} + (1-\alpha) {I_{i,t}^d}^{\phi}\right] + (1-\delta) \log H_{i,t-1} + \beta X_{i,t} + \varepsilon_{i,t}
$$

where $X_{i,t-1}$ contains demographic covariates, as well as mother characteristics. We assume that $\varepsilon_{i,t}$ follows a normal distribution with mean zero and variance σ_ε^2. The associated log-likelihood function is the following:

$$
\log \mathscr{L} = \sum_{i=1}^{n} \log\left[\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left(-\frac{\left(\log H_{i,t} - \frac{\nu}{\phi}\log\left[\alpha {I_{i,t}^e}^{\phi} + (1-\alpha){I_{i,t}^d}^{\phi}\right] - (1-\delta)\log H_{i,t-1} - \beta X_{i,t}\right)^2}{2\sigma_\varepsilon^2}\right)\right]
$$

$$= -\frac{n}{2}\log\left(\sigma_\varepsilon^2\right) - \frac{1}{2\sigma_\varepsilon^2}\sum_{i=1}^{n}\left(\log H_{i,t} - \frac{\nu}{\phi}\log\left[\alpha {I_{i,t}^e}^{\phi} + (1-\alpha){I_{i,t}^d}^{\phi}\right] - (1-\delta)\log H_{i,t-1} - \beta X_{i,t}\right)^2$$

The gradient of the likelihood function with respect to the parameter vector $\theta = (\sigma_\varepsilon^2, \nu, \phi, \alpha, \delta, \beta)$ is:

$$\frac{\partial \log \mathscr{L}}{\partial \sigma_\varepsilon^2} = -\frac{n}{2\sigma_\varepsilon^2} + \frac{1}{2\sigma_\varepsilon^4}\sum_{i=1}^{n}\left(\log H_{i,t} - \mu\right)^2$$

$$\frac{\partial \log \mathscr{L}}{\partial \nu} = \frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}\left(\log H_{i,t} - \mu\right)\frac{1}{\phi}\log\left[\alpha {I_{i,t}^e}^{\phi} + (1-\alpha){I_{i,t}^d}^{\phi}\right]$$

$$\begin{aligned}\frac{\partial \log \mathscr{L}}{\partial \phi} &= \frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}\left(\log H_{i,t} - \mu\right)\left\{-\frac{\nu}{\phi^2}\log\left[\alpha {I_{i,t}^e}^{\phi} + (1-\alpha){I_{i,t}^d}^{\phi}\right] + \right.\\ &\left. -\frac{\nu}{\phi\left[\alpha {I_{i,t}^e}^{\phi} + (1-\alpha){I_{i,t}^d}^{\phi}\right]}\alpha {I_{i,t}^e}^{\phi}\ln\left(I_{i,t}^e\right) + (1-\alpha){I_{i,t}^d}^{\phi}\ln\left(I_{i,t}^d\right)\right\}\\ &= \frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}\left(\log H_{i,t} - \mu\right)\frac{\nu}{\phi}\left\{-\frac{1}{\phi}\log\left(\Gamma\right) + \frac{1}{\Gamma}\left(\alpha {I_{i,t}^e}^{\phi}\ln\left(I_{i,t}^e\right) + (1-\alpha){I_{i,t}^d}^{\phi}\ln\left(I_{i,t}^d\right)\right)\right\}\end{aligned}$$

$$\frac{\partial \log \mathscr{L}}{\partial \alpha} = \frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}\left(\log H_{i,t} - \mu\right)\left\{\frac{\nu}{\phi}\frac{{I_{i,t}^e}^{\phi} - {I_{i,t}^d}^{\phi}}{\alpha {I_{i,t}^e}^{\phi} + (1-\alpha){I_{i,t}^d}^{\phi}}\right\}$$

$$\frac{\partial \log \mathscr{L}}{\partial (1-\delta)} = \frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}\left(\log H_{i,t} - \mu\right)\log H_{i,t-1}$$

$$\frac{\partial \log \mathscr{L}}{\partial \beta} = \frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}\left(\log H_{i,t} - \mu\right)X_{i,t}$$

where $\mu(\nu, \phi, \alpha, \delta, \beta) = \frac{\nu}{\phi}\log\left[\alpha {I_{i,t}^e}^{\phi} + (1-\alpha){I_{i,t}^d}^{\phi}\right] + (1-\delta)\log H_{i,t-1} + \beta X_{i,t}$ is the average of $\log H_{i,t}$;

$\Gamma(\phi, \alpha) = \alpha {I_{i,t}^e}^{\phi} + (1-\alpha){I_{i,t}^d}^{\phi}$ are the inside arguments of the CES function

$\mu_\nu^{'}(\nu, \phi, \alpha) = \frac{1}{\phi}\log\left(\Gamma\right)$ is the derivative of μ, or equivalently $\frac{\nu}{\phi}\log\left(\Gamma\right)$, with respect to ν;

$\mu_\phi^{'}(\nu, \phi, \alpha) = \frac{\nu}{\phi}\left\{\frac{\alpha {I_{i,t}^e}^{\phi}\ln\left(I_{i,t}^e\right) + (1-\alpha){I_{i,t}^d}^{\phi}\ln\left(I_{i,t}^d\right)}{\Gamma} - \frac{1}{\phi}\log\left(\Gamma\right)\right\}$ is the derivative of μ with respect to ϕ;

$\mu_\alpha^{'}(\nu, \phi, \alpha) = \frac{\nu}{\phi}\left\{\frac{{I_{i,t}^e}^{\phi} - {I_{i,t}^d}^{\phi}}{\Gamma}\right\}$ is the derivative of μ with respect to α;

In order to appropriately estimate the error terms, the Hessian must be also calculated.

The Hessian of the likelihood function with respect to the parameter σ_ε^2 is:

$$
\begin{aligned}
\frac{\partial \log \mathscr{L}}{\partial \sigma_\varepsilon^2 \partial \sigma_\varepsilon^2} &= \frac{n}{2\sigma_\varepsilon^4} - \frac{1}{\sigma_\varepsilon^6} \sum_{i=1}^{n} (\log H_{i,t} - \mu)^2 \\
\frac{\partial \log \mathscr{L}}{\partial \sigma_\varepsilon^2 \partial \nu} &= -\frac{1}{\sigma_\varepsilon^4} \sum_{i=1}^{n} (\log H_{i,t} - \mu)\, \mu_\nu^{'}(\nu, \phi, \alpha) \\
\frac{\partial \log \mathscr{L}}{\partial \sigma_\varepsilon^2 \partial \phi} &= -\frac{1}{\sigma_\varepsilon^4} \sum_{i=1}^{n} (\log H_{i,t} - \mu)\, \mu_\phi^{'}(\nu, \phi, \alpha) \\
\frac{\partial \log \mathscr{L}}{\partial \sigma_\varepsilon^2 \partial \alpha} &= -\frac{1}{\sigma_\varepsilon^4} \sum_{i=1}^{n} (\log H_{i,t} - \mu)\, \mu_\alpha^{'}(\nu, \phi, \alpha) \\
\frac{\partial \log \mathscr{L}}{\partial \sigma_\varepsilon^2 \partial (1-\delta)} &= -\frac{1}{\sigma_\varepsilon^4} \sum_{i=1}^{n} (\log H_{i,t} - \mu) \log H_{i,t-1} \\
\frac{\partial \log \mathscr{L}}{\partial \sigma_\varepsilon^2 \partial \beta} &= -\frac{1}{\sigma_\varepsilon^4} \sum_{i=1}^{n} (\log H_{i,t} - \mu)\, X_{i,t}
\end{aligned}
$$

Taking expectations we have that $E\left[\sum_{i=1}^{n} (\log H_{i,t} - \mu)^2\right] = E[\epsilon'\epsilon] = n\sigma_\varepsilon^2$ and that $E\left[\sum_{i=1}^{n} (\log H_{i,t} - \mu)\, X\right] = 0$. Therefore we have that all the cross-partials are zero, and the only relevant second derivative boils down to

$$
\frac{\partial \log \mathscr{L}}{\partial \sigma_\varepsilon^2 \partial \sigma_\varepsilon^2} = \frac{n}{2\sigma_\varepsilon^4} - \frac{n\sigma_\varepsilon^2}{\sigma_\varepsilon^6} = -\frac{n}{2\sigma_\varepsilon^4}
$$

It can be shown that the expected value of the Hessian of the likelihood function with respect to the rest of the parameters $\theta = (\nu, \phi, \alpha, 1-\delta, \beta)$ is equal to $-1/\sigma_\varepsilon^2$ times the cross product of the following matrix:

$$
\Omega = \left[\mu_\nu^{'}(\nu, \phi, \alpha); \mu_\phi^{'}(\nu, \phi, \alpha); \mu_\alpha^{'}(\nu, \phi, \alpha); \log H_{i,t-1}; X_{i,t}\right]
$$

.

Therefore we have that

$$
E\left(H\right) = -\begin{bmatrix} \frac{n}{2\sigma_\varepsilon^4} & 0 \\ 0 & \frac{\Omega'\Omega}{\sigma_\varepsilon^2} \end{bmatrix}
$$

The tedious derivation can be found here below.

The Hessian of the likelihood function with respect to the parameter ν is :

$$
\begin{aligned}
\frac{\partial \log \mathscr{L}}{\partial \nu \partial \nu} &= -\frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^{n} \left[\mu_\nu^{'}(\nu, \phi, \alpha)\right]^2 \\
\frac{\partial \log \mathscr{L}}{\partial \nu \partial \phi} &= -\frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^{n} \mu_\nu^{'}(\nu, \phi, \alpha) \mu_\phi^{'}(\nu, \phi, \alpha) - (\log H_{i,t} - \mu)\, \frac{1}{\nu} \mu_\phi^{'}(\nu, \phi, \alpha)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \log \mathscr{L}}{\partial \nu \partial \alpha} &= -\frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n} \mu_\nu^{'}(\nu,\phi,\alpha)\mu_\alpha^{'}(\nu,\phi,\alpha) - (\log H_{i,t} - \mu)\,\frac{1}{\nu}\mu_\alpha^{'}(\nu,\phi,\alpha) \\
\frac{\partial \log \mathscr{L}}{\partial \nu \partial (1-\delta)} &= -\frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n} \mu_\nu^{'}(\nu,\phi,\alpha)\log H_{i,t-1} \\
\frac{\partial \log \mathscr{L}}{\partial \nu \partial \beta} &= -\frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n} \mu_\nu^{'}(\nu,\phi,\alpha)X_{i,t}
\end{aligned}
$$

Taking expectations we find that the terms $\mu_\phi^{'}(\nu,\phi,\alpha)$ and $\mu_\alpha^{'}(\nu,\phi,\alpha)$ that are multiplied by the error term $\sum_{i=1}^{n}(\log H_{i,t} - \mu)$ drop out, since $E(\epsilon) = 0$.

The Hessian of the likelihood function with respect to the parameter ϕ is :

$$
\begin{aligned}
\frac{\partial \log \mathscr{L}}{\partial \phi \partial \phi} &= \frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}(\log H_{i,t} - \mu)\,\frac{-\mu_\phi^{'}(\nu,\phi,\alpha)}{\phi} - \left[\mu_\phi^{'}(\nu,\phi,\alpha)\right]^2 + \\
&\quad + (\log H_{i,t} - \mu)\,\frac{\nu}{\phi}\left[\frac{\log(\Gamma)}{\phi^2} - \frac{1}{\phi}\frac{\alpha {I_{i,t}^e}^{\phi}\ln\left(I_{i,t}^e\right) + (1-\alpha){I_{i,t}^d}^{\phi}\ln\left(I_{i,t}^d\right)}{\Gamma}\right] \\
&\quad + (\log H_{i,t} - \mu)\,\frac{\nu}{\phi}\left[\frac{\alpha {I_{i,t}^e}^{\phi}\ln\left(I_{i,t}^e\right)^2 + (1-\alpha){I_{i,t}^d}^{\phi}\ln\left(I_{i,t}^d\right)^2}{\Gamma} + \right. \\
&\quad \left. - \left(\frac{\alpha {I_{i,t}^e}^{\phi}\ln\left(I_{i,t}^e\right) + (1-\alpha){I_{i,t}^d}^{\phi}\ln\left(I_{i,t}^d\right)}{\Gamma}\right)^2\right] \\
&= -\frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}\left[\mu_\phi^{'}(\nu,\phi,\alpha)\right]^2 + (\log H_{i,t} - \mu)\,\frac{\mu_\phi^{'}(\nu,\phi,\alpha)}{\phi} + \\
&\quad + (\log H_{i,t} - \mu)\,\frac{\mu_\phi^{'}(\nu,\phi,\alpha)}{\phi} + \\
&\quad - (\log H_{i,t} - \mu)\,\frac{\nu}{\phi}\left[\frac{\alpha {I_{i,t}^e}^{\phi}\ln\left(I_{i,t}^e\right)^2 + (1-\alpha){I_{i,t}^d}^{\phi}\ln\left(I_{i,t}^d\right)^2}{\Gamma} + \right. \\
&\quad \left. - \left(\frac{\alpha {I_{i,t}^e}^{\phi}\ln\left(I_{i,t}^e\right) + (1-\alpha){I_{i,t}^d}^{\phi}\ln\left(I_{i,t}^d\right)}{\Gamma}\right)^2\right] \\
&= -\frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}\left[\mu_\phi^{'}(\nu,\phi,\alpha)\right]^2 + (\log H_{i,t} - \mu)\,A(\nu,\phi,\alpha) \\
\frac{\partial \log \mathscr{L}}{\partial \phi \partial \alpha} &= -\frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}\mu_\phi^{'}(\nu,\phi,\alpha)\mu_\alpha^{'}(\nu,\phi,\alpha) + (\log H_{i,t} - \mu)\,\mu_{\phi\alpha}^{''}(\nu,\phi,\alpha)
\end{aligned}
$$

$$\begin{aligned}\frac{\partial \log \mathscr{L}}{\partial\phi\partial(1-\delta)} &= -\frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}\mu_\phi^{'}(\nu,\phi,\alpha)\log H_{i,t-1}\\ \frac{\partial \log \mathscr{L}}{\partial\phi\partial\beta} &= -\frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}\mu_\phi^{'}(\nu,\phi,\alpha)X_{i,t}\end{aligned}$$

Taking expectations, all the constant terms multiplied by $\sum_{i=1}^{n}(\log H_{i,t}-\mu)$ drop out, since $E(\epsilon)=0$.[51] Therefore we have that on the diagonal we have the usual squared term: $\frac{\partial \log \mathscr{L}}{\partial\phi\partial\phi} = -\frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}\left[\mu_\phi^{'}(\nu,\phi,\alpha)\right]^2$

The Hessian of the likelihood function with respect to the parameter α is :

$$\begin{aligned}\frac{\partial \log \mathscr{L}}{\partial^2\alpha} &= -\frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}\left[\mu_\alpha^{'}(\nu,\phi,\alpha)\right]^2-(\log H_{i,t}-\mu)\,\frac{\nu}{\phi}\left[\frac{{I_{i,t}^e}^{\phi}-{I_{i,t}^d}^{\phi}}{\alpha{I_{i,t}^e}^{\phi}+(1-\alpha){I_{i,t}^d}^{\phi}}\right]^2\\ \frac{\partial \log \mathscr{L}}{\partial\alpha\partial(1-\delta)} &= -\frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}\mu_\alpha^{'}(\nu,\phi,\alpha)\log H_{i,t-1}\\ \frac{\partial \log \mathscr{L}}{\partial\alpha\partial\beta} &= -\frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}\mu_\alpha^{'}(\nu,\phi,\alpha)X_{i,t}\end{aligned}$$

Again, taking expectations all the terms multiplied by ϵ drop out, so that we are left with $\frac{\partial \log \mathscr{L}}{\partial^2\alpha} = -\frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}\left[\mu_\alpha^{'}(\nu,\phi,\alpha)\right]^2$.

The Hessian of the likelihood function with respect to the parameter δ is :

$$\begin{aligned}\frac{\partial \log \mathscr{L}}{\partial^2(1-\delta)} &= -\frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}(\log H_{i,t-1})^2\\ \frac{\partial \log \mathscr{L}}{\partial\beta\partial(1-\delta)} &= -\frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}\log H_{i,t-1}X_{i,t}\end{aligned}$$

And finally, the Hessian of the likelihood function with respect to the parameter β is the usual:

$$\frac{\partial \log \mathscr{L}}{\partial^2\beta} = -\frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{n}X_{i,t}^2$$

If we allow some of the parameters to change according to the genotype $g_{i,t}$ of the individual we have the following log-likelihood:

[51] The constant term $A(\nu,\phi,\alpha)$ is equal to $\left\{2\frac{\mu_\phi^{'}(\nu,\phi,\alpha)}{\phi}+\frac{\nu}{\phi\Gamma}\left[\left(\alpha{I_{i,t}^e}^{\phi}\ln I_{i,t}^e+(1-\alpha){I_{i,t}^d}^{\phi}\ln I_{i,t}^d\right)^2/\Gamma\right.\right.$
$\left.\left.-\alpha{I_{i,t}^e}^{\phi}\ln\left(I_{i,t}^e\right)^2-(1-\alpha){I_{i,t}^d}^{\phi}\ln\left(I_{i,t}^d\right)^2\right]\right\}$

$$\log \mathscr{L} = -\frac{n}{2}\log\left(\sigma_{\varepsilon}^{2}\right) - \frac{1}{2\sigma_{\varepsilon}^{2}}\sum_{i=1}^{n}\left(\log H_{i,t} - \frac{\nu}{\phi_0+\phi_1 g_{i,t}}\log\left[\Delta\right] - (1-\delta_0+\delta_1 g_{i,t})\log H_{i,t-1} - \beta X_{i,t}\right)^{2}$$

Where $\Delta = (\alpha+\alpha_1 g_{i,t}){I_{i,t}^{e}}^{\phi_0+\phi_1 g_{i,t}} + (1-\alpha_0-\alpha g_{i,t}){I_{i,t}^{d}}^{\phi_0+\phi_1 g_{i,t}}$. The gradient with respect to additional the parameters $(\nu_1, \phi_1, \alpha_1, \delta_1)$ is the same as the ones above, simply multiplied by $g_{i,t}$:

$$\begin{aligned}
\frac{\partial \log \mathscr{L}}{\partial \nu_1} &= \frac{1}{\sigma_{\varepsilon}^{2}}\sum_{i=1}^{n}\left(\log H_{i,t}-\mu\right)\frac{1}{\phi}\log\left[\Delta\right]\times g_{i,t} \\
\frac{\partial \log \mathscr{L}}{\partial \phi_1} &= \frac{1}{\sigma_{\varepsilon}^{2}}\sum_{i=1}^{n}\left(\log H_{i,t}-\mu\right)\left\{-\frac{\nu}{\left(\phi_0+\phi_1 g_{i,t}\right)^{2}}g_{i,t}\log\left[\Delta\right] + \frac{\nu}{\phi_0+\phi_1 g_{i,t}}\frac{1}{\Delta}\left(\phi_0+\phi_1 g_{i,t}\right)\cdot\right. \\
&\quad \left.\cdot\, g_{i,t}\left((\alpha+\alpha_1 g_{i,t}){I_{i,t}^{e}}^{\phi_0+\phi_1 g_{i,t}} + (1-\alpha_0-\alpha g_{i,t}){I_{i,t}^{d}}^{\phi_0+\phi_1 g_{i,t}}\right)\right\} \\
&= \frac{1}{\sigma_{\varepsilon}^{2}}\sum_{i=1}^{n}\left(\log H_{i,t}-\mu\right)\left\{-\frac{\nu}{\left(\phi_0+\phi_1 g_{i,t}\right)^{2}}\log\left[\Delta\right] + \frac{\nu}{\phi_0+\phi_1 g_{i,t}}\frac{1}{\Delta}\left(\phi_0+\phi_1 g_{i,t}\right)\cdot\right. \\
&\quad \left.\cdot\left((\alpha+\alpha_1 g_{i,t}){I_{i,t}^{e}}^{\phi_0+\phi_1 g_{i,t}} + (1-\alpha_0-\alpha g_{i,t}){I_{i,t}^{d}}^{\phi_0+\phi_1 g_{i,t}}\right)\right\}\times g_{i,t} \\
\frac{\partial \log \mathscr{L}}{\partial (1-\delta)} &= \frac{1}{\sigma_{\varepsilon}^{2}}\sum_{i=1}^{n}\left(\log H_{i,t}-\mu\right)\left(-\log H_{i,t-1}\right)\times g_{i,t}
\end{aligned}$$

C Summary Statistics and Distributions

C.1 Summary Statistics

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

Table (10) 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.[52]. 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

[52]See Cole et al. (1998)

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 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(11) reports the average value of the controls 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 10: Summary Statistics

	FTO gene type		
	T-Allele	A-Risky	Total
Height	142.71	143.03	142.91
(cm)	[0.22]	[0.17]	[0.13]
Weight	39.19***	40.01***	39.71
(kg)	[0.18]	[0.14]	[0.11]
BMI	18.20***	18.50***	18.39
	[0.04]	[0.03]	[0.02]
BMI z-score	0.17***	0.27***	0.24
	[0.01]	[0.01]	[0.01]
Fat Percentage	24.10***	25.25***	24.83
	[0.14]	[0.11]	[0.09]
Overweight (%)	19.91***	24.58***	22.87
	[0.42]	[0.34]	[0.27]
Underweight (%)	3.89	4.00	3.96
	[0.20]	[0.16]	[0.12]
Kilocalories	1.82***	1.84***	1.83
(x1000)	[0.01]	[0.00]	[0.00]
Fat Intake	73.05***	74.20***	73.78
(grams/day)	[0.28]	[0.21]	[0.17]
Dietary Cholesterol Intake	181.24*	183.88*	182.91
(grams/day)	[1.20]	[0.91]	[0.73]
Carbohydrate Intake	243.50***	246.57***	245.45
(grams/day)	[0.82]	[0.60]	[0.49]
Total Sugar Intake	111.89**	113.54**	112.94
(grams/day)	[0.58]	[0.41]	[0.33]
Physical Activity	7.69	7.72	7.71
(Sedentary Hours)	[0.02]	[0.02]	[0.01]
Physical Activity	23.78	23.58	23.65
(Moderate To Vigorous)	[0.28]	[0.21]	[0.17]
Physical Activity	564.15	560.37	561.74
(counts per minute)	[3.22]	[2.43]	[1.94]
Very Active	3.44	3.47	3.46
(self-report)	[0.03]	[0.02]	[0.02]

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 (kg/m^2) normal z-scores calculated using 1990 British Growth Reference. Fat percentage: ratio of fat mass to total mass, calculated using Lunar Prodigy DXA scanner. 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 minute, minutes of sedentary activity, and minutes of moderate-to-vigorous activity (MVPA; more than 3600 counts/min, standard cutoff). Self-reported participation in physical activity ranged from 1 (never) to 5 (daily).

Table 11: Control Variables, by Child FTO genotype

	FTO gene type		
	T-Allele	A-Risky	Total
Mother Edu	3.15	3.14	3.14
	[0.02]	[0.02]	[0.01]
Father Edu	3.18	3.18	3.18
	[0.03]	[0.02]	[0.02]
Mother SES	2.81	2.84	2.82
	[0.02]	[0.02]	[0.01]
Father SES	2.95	2.93	2.94
	[0.03]	[0.02]	[0.02]
Mother BMI	22.79**	23.03**	22.94
	[0.07]	[0.06]	[0.05]
Mother age	28.55	28.62	28.59
at birth	[0.09]	[0.06]	[0.05]
Teen mother	0.05	0.04	0.04
	[0.00]	[0.00]	[0.00]
Single Mother	0.19	0.19	0.19
	[0.01]	[0.01]	[0.00]
Parity	0.74	0.76	0.75
	[0.02]	[0.01]	[0.01]
Birth Weight	3442.69	3450.44	3447.59
(gr)	[9.29]	[7.13]	[5.66]

Average value of the controls used in the 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 7: Distribution of BMI, Females

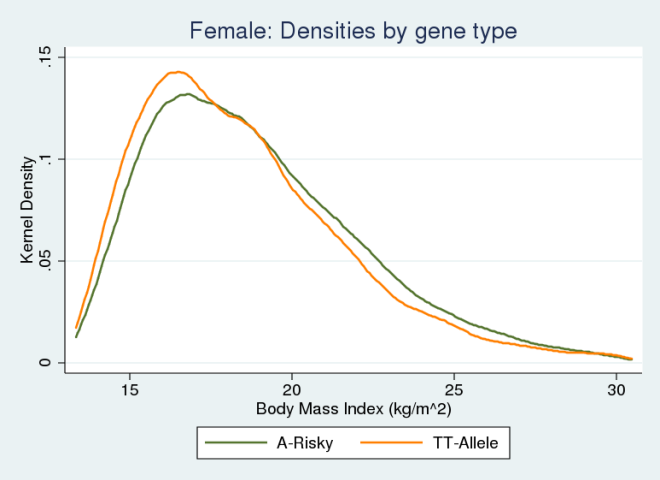
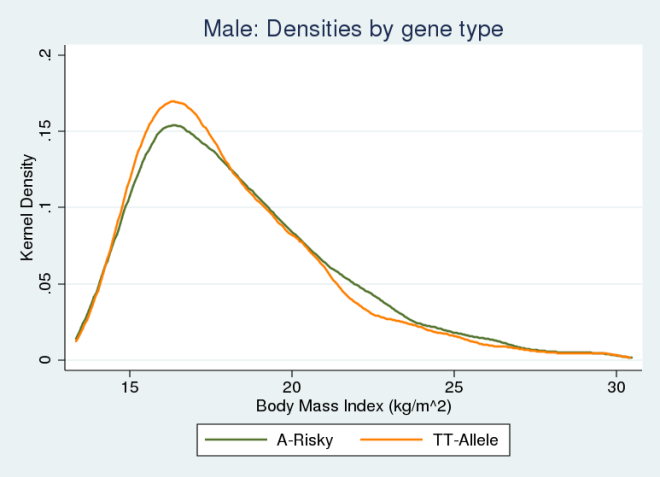


Figure 8: Distribution of BMI, Males



C.2 Distributions

Here below are the empirical distributions of the relevant variables, divided by gender and genetic endowment g. Figures (7) and (8) display the distribution of Body Mass Index, H; figures (9) and (10) display the distribution of the investment in diet, I_d; figures (11) and (12) display the distribution of the investment in exercise, I_e;

D Robustness Checks

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

Tables (12) and (13) report the estimation of the linear health production separately by gender. Table (14) reports the estimates for the robustness checks, using the genetic

Figure 9: Distribution of Diet, Females

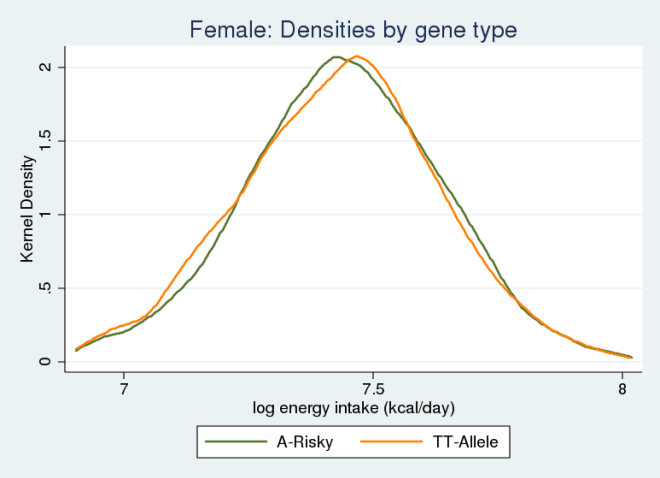


Figure 10: Distribution of Diet, Males

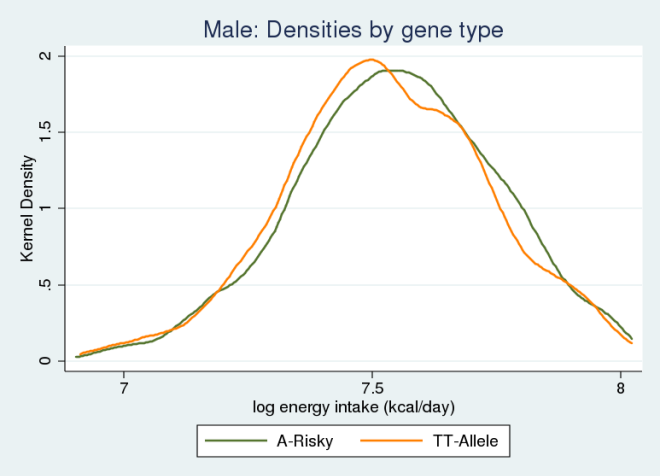


Figure 11: Distribution of Activity, Females

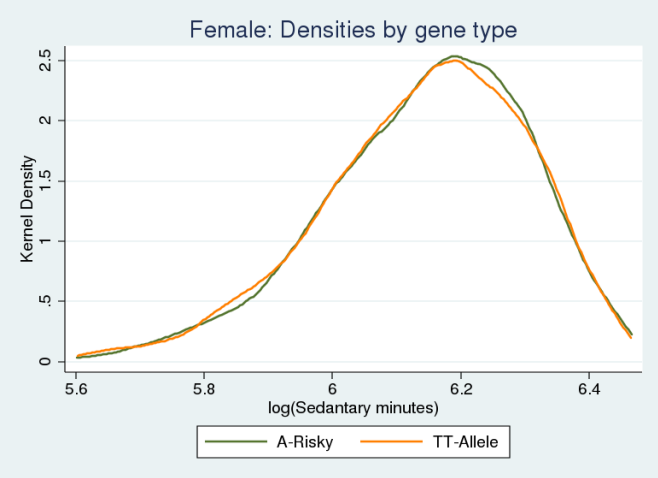
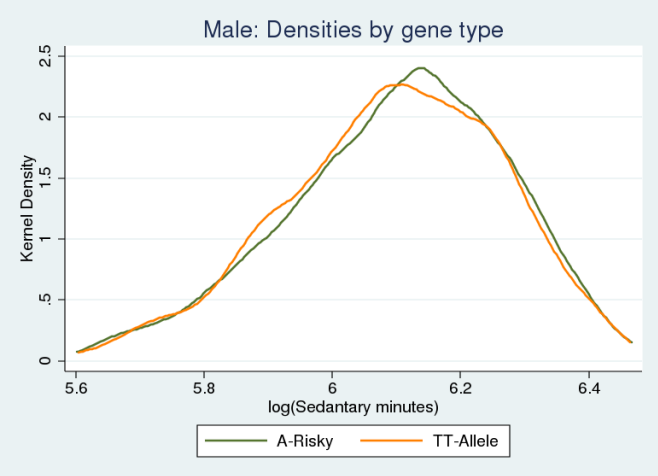


Figure 12: Distribution of Activity, Males



score as a measure of g. Tables (15) and (16) report the estimation of the linear health production function using different measures of investments, both for diet I_d and exercise I_e.

Table 12: Gene and Investment Interaction - FTO, Females

		(1)	(2)	(3)	(4)	(5)
Risky FTO Gene	β_g	0.022	0.003	0.006	0.009	0.012
		[0.007]***	[0.003]	[0.003]*	[0.003]***	[0.003]***
log(Energy Intake)	α_d			0.054	0.083	0.082
				[0.009]***	[0.014]***	[0.014]***
G X Energy Intake	$\alpha_{g\times d}$				0.043	0.047
					[0.018]**	[0.018]***
log(Sedentary min.)	α_e			0.010	0.012	0.005
				[0.007]	[0.013]	[0.013]
G X Sedentary min.	$\alpha_{g\times e}$				-0.006	-0.002
					[0.016]	[0.016]
Mom Risky Gene	γ_g		-0.006	-0.006	-0.006	
			[0.004]*	[0.004]	[0.004]	
log(BMI_{mom})	γ_h		0.104	0.104	0.103	
			[0.010]***	[0.010]***	[0.010]***	
$log(BMI)_{t-1}$	$(1-\delta)$		0.950	0.929	0.929	0.961
			[0.010]***	[0.011]***	[0.011]***	[0.011]***
Controls			X	X	X	
R^2		0.41%	78%	78%	78%	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); Controls: 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; late respondent; birth weight.

Table 13: 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]**	[0.005]	[0.005]
log(Energy Intake)	α_d			0.065	0.067	0.069
				[0.008]***	[0.013]***	[0.013]***
G X Energy Intake	$\alpha_{g\times d}$				0.003	0.004
					[0.016]	[0.016]
log(Sedentary min.)	α_e			0.010	0.042	0.038
				[0.007]	[0.013]***	[0.013]***
G X Sedentary min.	$\alpha_{g\times e}$				0.027	0.023
					[0.016]*	[0.016]
Mom Risky Gene	γ_g		0.000	0.001	0.001	
			[0.004]	[0.004]	[0.004]	
$\log(BMI_{mom})$	γ_h		0.073	0.076	0.076	
			[0.011]***	[0.011]***	[0.011]***	
$log(BMI)_{t-1}$	$(1-\delta)$		0.991	0.947	0.947	0.972
			[0.011]***	[0.012]***	[0.012]***	[0.011]***
Controls			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); Controls: 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; late respondent; birth weight.

Table 14: Robustness Checks - Genetic Score

		(1) Baseline	(2) Males	(3) Females	(4) No Underweight	(5) RE	(6) FE	(7) Prob Overweight	(8) Weight	(9) zBMI	(10) Fat %
Risky Genetic Score	β_g	0.013 [0.003]***	0.012 [0.003]***	0.006 [0.003]**	0.008 [0.002]***	0.012 [0.003]***		0.073 [0.046]	0.011 [0.002]***	0.068 [0.015]***	0.026 [0.014]*
log(Energy Int.)	α_d	0.065 [0.008]***	0.063 [0.010]***	0.064 [0.011]***	0.053 [0.007]***	0.050 [0.008]***	0.016 [0.009]*	0.456 [0.168]***	0.054 [0.009]***	0.403 [0.052]***	-0.012 [0.057]
G X Energy Int.	$\alpha_{g\times d}$	0.007 [0.008]	-0.003 [0.012]	0.019 [0.012]	0.007 [0.008]	0.010 [0.008]	-0.002 [0.010]	0.032 [0.169]	0.003 [0.009]	0.074 [0.055]	-0.047 [0.057]
log(Sedentary m.)	α_e	0.021 [0.008]**	0.021 [0.012]*	0.019 [0.012]*	0.023 [0.008]***	0.033 [0.009]***	0.029 [0.011]***	0.539 [0.195]***	0.029 [0.009]***	0.148 [0.060]**	0.039 [0.058]
G X Sedentary m.	$\alpha_{g\times e}$	0.000 [0.011]	-0.008 [0.015]	0.010 [0.015]	0.000 [0.011]	-0.005 [0.011]	-0.012 [0.014]	0.007 [0.242]	0.008 [0.012]	-0.001 [0.077]	-0.078 [0.079]
H_{t-1}	$(1-\delta)$	0.938 [0.008]***	0.945 [0.012]***	0.929 [0.011]***	0.911 [0.008]***	0.813 [0.009]***	-0.135 [0.017]***	2.092 [0.052]***	0.759 [0.008]***	0.868 [0.008]***	0.309 [0.023]***
Controls		X	X	X	X	X	X	X	X	X	X
R^2		0.78	0.79	0.78	0.77		0.64		0.88	0.77	0.54
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 5). 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. Controls: 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; late respondent; birth weight.

Table 15: Different Measures of Diet and Food Intake - FTO gene

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Calories	Proteins	Fat	Carbs	Dietary Cholesterol	Sugar	Starch	Non Starch	Factor Score
Risky FTO Gene	β_g	0.009	0.009	0.008	0.008	0.007	0.006	0.007	0.008	0.008
		[0.003]***	[0.002]***	[0.002]***	[0.003]***	[0.002]***	[0.002]***	[0.002]***	[0.002]***	[0.002]***
log(Diet)	α_d	0.068	0.046	0.038	0.047	0.01	0.011	0.046	0.023	0.016
		[0.009]***	[0.007]***	[0.007]***	[0.008]***	[0.004]***	[0.005]**	[0.007]***	[0.005]***	[0.002]***
G X Diet	$\alpha_{g\times d}$	0.024	0.026	0.014	0.012	0.008	0.002	0.01	0.014	0.005
		[0.011]**	[0.009]***	[0.008]*	[0.010]	[0.005]*	[0.006]	[0.009]	[0.007]**	[0.003]*
log(Sedentary min.)	α_e	0.029	0.028	0.029	0.028	0.026	0.026	0.029	0.027	0.03
		[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	0.01	0.013	0.011	0.01	0.01	0.012	0.011	0.013
		[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]	[0.011]
Mom Risky Gene	γ_g	-0.003	-0.003	-0.002	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003
		[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
$log(BMI)_{t-1}$	$(1-\delta)$	0.94	0.941	0.945	0.943	0.947	0.947	0.944	0.948	0.938
		[0.008]***	[0.008]***	[0.008]***	[0.008]***	[0.008]***	[0.008]***	[0.008]***	[0.008]***	[0.008]***
Controls		X	X	X	X	X	X	X	X	X
R^2		0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78
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: energy 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); Controls: 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; late respondent; birth weight.

Table 16: Different Measures of Physical Activity - FTO gene

		(1) Sedentary min	(2) MVPA	(3) Counts per min	(4) Factor Score
Risky FTO Gene	β_g	0.009	0.008	0.009	0.009
		[0.003]***	[0.003]***	[0.003]***	[0.003]***
log(Energy Intake)	α_d	0.068	0.068	0.069	0.069
		[0.009]***	[0.009]***	[0.009]***	[0.009]***
G X Energy Intake	$\alpha_{g\times d}$	0.024	0.02	0.023	0.021
		[0.011]**	[0.011]*	[0.011]**	[0.011]*
log(Exercise)	α_e	0.029	-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]
Mom Risky Gene	γ_g	-0.003	-0.003	-0.003	-0.003
		[0.003]	[0.003]	[0.003]	[0.003]
$log(BMI)_{t-1}$	$(1-\delta)$	0.94	0.935	0.937	0.937
		[0.008]***	[0.008]***	[0.008]***	[0.008]***
Controls		X	X	X	X
R^2		0.78	0.79	0.78	0.78
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); Controls: 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; late respondent; birth weight.